A pilot study to determine the relative value of non-market transport impacts of investment

December 2018

Tim Denne[1], Geoff Kerr[2], David Glover,[3] Michael Winder[3] and Louis Wright[1]
[1]Resource Economics, [2]Lincoln University, [3]Gravitas Research & Strategy

NZ Transport Agency research report 648
Contracted research organisation – Resource Economics

ISBN 978-1-98-856113-4 (electronic)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Denne, T, G Kerr, D Glover, M Winder and L Wright (2018) A pilot study to determine the relative value of non-market transport impacts of investment. *NZ Transport Agency research report 648.* 121pp.

Resource Economics was contracted by the NZ Transport Agency in 2016 to carry out this research.

Keywords: choice modelling, non-market valuation, reliability, survey design, value of statistical life, value of time

An important note for the reader

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an efficient, effective and safe land transport system in the public interest. Each year, the NZ Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in research reports are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the NZ Transport Agency. The material contained in the reports should not be construed in any way as policy adopted by the NZ Transport Agency or indeed any agency of the NZ Government. The reports may, however, be used by NZ Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation, the NZ Transport Agency and agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

Acknowledgements

We are grateful to the many people who have assisted with this project, particularly our steering group (Joanne Leung, Jennifer McSaveney, Sandy Fong, Paul Graham and Peter King) and our peer reviewers: Dr Richard Meade (Cognitus Economic Insight) and Dr Adolf Stroombergen (Infometrics).

We are also extremely grateful to the input and advice to the project provided by Professor John Rose and Professor Riccardo Scarpa.

Contents

Executive summary

Study objectives

This research project examined methodologies within a single survey setting to establish the relative monetary value of selected transport impacts which cannot be measured using market values. These include the costs of travel time and the value of reducing the risk of injuries or fatalities. Monetary values are included in the *Economic evaluation manual* (EEM) (NZ Transport Agency 2016), but survey and analysis methodologies have improved, and preferences may have changed over time. This study examined whether a new survey, using best practice methods, could be used to obtain robust values for more than one non-market transport impact.

Non-market values

Non-market values are used as inputs to decisions that weigh up the costs and benefits of investment and operational decisions. For example, how to trade-off time savings for safety, or the willingness to pay (WTP) for additional time savings or safety enhancements.

There are many potential factors that might be weighed up, but the emphasis in this study was on the following.

- Value of statistical life – measures taken to reduce the number of fatal crashes do not eliminate fatalities but reduce the probability or risk of a fatality. Approaches to valuation estimate the value to society of incremental changes in the risk of death. These values are referred to as the value of statistical life (VoSL or VSL).
- Value of prevented injuries – the value to road users of reductions in the risk of non-fatal injuries.
- Value of travel time savings – the value to road users of reducing the travel time of a journey.
- Value of reliability – the value to road users of reducing the variability in travel time.
- Value of reduced congestion – the value to road users of reducing the time within a trip spent in heavy traffic.

Choice modelling

Choice modelling (CM) and contingent valuation (CV) are stated preference (SP) methods commonly used to infer non-market values. CV has been the traditional SP tool for non-market valuation, but choice experiments provide more information on consumer preferences than is possible through CV. CM is one of the methodologies investigated in this study.

CM techniques quantify the trade-offs people make between specific attributes of a good or service. This is done by presenting a series of hypothetical scenarios in which respondents select their preferred choice among a number of options, otherwise known as a choice set. Reliable choice surveys or choice experiments simulate real-world scenarios to the extent possible and include all key attributes in the description of the good or service. This study applied CM in scenarios that required participants to choose between two alternative routes. The routes differed in the attributes of interest, with the differences carefully designed based on statistical criteria. When respondents chose their preferred route for several choice-pairs, the data could then be analysed econometrically to understand the strength of the factors contributing to these relative preferences.

For valuation purposes, a monetised attribute (cost) was included so dollar values of non-market goods could be estimated from the trade-off between, eg safety and cost.

Recent international studies have surveyed respondents, asking them to choose between two routes for a journey, one of which allows, eg faster travel (it has higher speed limits), but has a longer total journey time, higher fuel costs, and a greater risk of death or injury. Results are expressed as WTP per trip. To convert WTP per trip to a VoSL, or a value of reducing risk of a crash, WTP values have to be converted into a value per person per kilometre, which is then multiplied by the inverse of the chance of death or injury per kilometre, ie the numbers of fatalities and injuries in each injury class, divided by total vehicle kilometres travelled (VKT).

Building a choice experiment for New Zealand

The study examined whether CM could be used to derive values for several transport outcomes simultaneously. SP study designs must weigh up the desire to obtain information about multiple values with the added complexity for respondents (and for the analysis) of adding more dimensions to a choice experiment. Table E.1 describes the design dimensions central to SP surveys.

Table E.1 Dimensions of a SP study

Dimension	Description
Choice tasks	The number of choice questions offered, eg one choice question would be ‘Do you prefer route A or route B?’
Alternatives	The number of options to select from in each choice task. In this study, each choice offered two alternatives (route A or B).
Attributes	The number of characteristics for each alternative, eg average travel time, lateness, congestion
Attribute levels	Variations in the measure of an attribute offered across the alternatives, eg average trip travel times of 20 minutes, 35 minutes, 40 minutes etc.

We built on recent international studies that have used CM for transport applications, but examined whether additional elements might be added to enable the valuation of more attributes. The process of development included two phases of initial testing with potential respondents, followed by a small pilot test.

Testing

Two initial phases of testing were used to obtain feedback on the design of the choice experiment, the number of attributes and the presentation of the alternatives. Following feedback from these two phases, the number of attributes was reduced and the presentation of the choices was simplified; this included dropping an experiment with using a simplified map (similar to a GPS format), because it appeared people were taking account of information on the map in a way which could not be measured. These changes were the basis for a pilot test.

Pilot test

The pilot test used a simple presentation of the choice tasks. It included the following attributes: 1) average travel time; 2) lateness – the percentage of trips delayed by a specified amount; 3) the percentage of the trip in heavy traffic; 4) the cost of the trip; 5) the number of deaths, serious and minor injuries on the route per billion kilometres travelled. The latter was used as a proxy for risk of a fatality or injury from future travel using these routes.

Based on earlier phase results, an experimental design for the pilot survey was developed to maximise statistical significance of the results. This included defining the number of choice tasks each respondent was given (10 were used) and the number of variants of the choice tasks (six in total). It also defined the combination of attribute levels across the individual attributes within each choice set.

The pilot test obtained 72 responses, including 22 completed face-to-face at a central Auckland location and 50 completed online by respondents around the rest of New Zealand. Respondents included a mix of gender, ethnicities, ages and socio-economic backgrounds.

Results and conclusions

Several analytical models were used to analyse the data collected. All produced results which permitted estimation of WTP for changes in attributes and VoSL, the value of injuries, the value of time and of lateness. Although value estimates have been produced (described in section 7.3), the small survey size means the error margins are large. The values should not be used to draw conclusions about the mean values which might be derived from a future, larger survey. We have not included the results in this summary because of the levels of uncertainty.

This study has confirmed it is possible to use a CM survey to estimate VoSL, the value of injuries and the value of time. This suggests it is worthwhile proceeding with the design and implementation of a full survey with more participants. The survey described in this report, with the suggested changes, would be a suitable basis for such a survey and analysis.

Abstract

This research project examined whether a single new survey, and analysis of the data, could be used to obtain robust values for the monetary value of statistical life, prevented injuries, travel time savings, trip reliability and congestion. A review of approaches used elsewhere was followed by the design of a choice modelling survey and two rounds of initial testing. This was used to develop a pilot survey with an efficient experimental design, and implemented online and face-to-face with 72 people. The data was analysed to produce statistically significant values for all parameters. This suggests it is worthwhile proceeding with the design and implementation of a full survey with more participants. The survey described in this report, with the suggested changes, would be a suitable basis for such a survey and analysis.

1 Introduction

1.1 Objectives of this project

This research project examined methodologies to improve current estimates of the monetary value of selected transport impacts which cannot be measured using market values, including the costs of travel time and the risk of injuries or fatalities.

Currently several non-market values are included in the *Economic evaluation manual* (EEM) (NZ Transport Agency 2016) based on historical surveys, with values updated annually using cost indexes. However, many of the values have not been re-analysed for several years, eg the value placed on reducing road fatalities is based on a survey from 27 years ago. Survey and analysis methodologies have improved, and preferences may have changed over time. This study examined whether a new survey, using best practice methods, could obtain robust values for more than one non-market transport impact. This study is not that new survey; rather, it has examined the options for a new survey and analysis.

The research built on previous research by the NZ Institute of Economic Research (NZIER) (Clough et al 2015) which provided guidance on approaches to valuing injury and mortality risks distilled from international literature. The authors noted the estimates used in New Zealand studies for the value of reducing the risk of fatal crashes had been updated to take account of changes in income, but an additional significant factor was the base level of risk. The number of road deaths in New Zealand has reduced from 21.4 per 100,000 population in 1990 to 6.9 per 100,000 population in 2015 (Ministry of Transport 2016a). The authors suggested a new survey to update the current estimates of the value of reducing the risks of crashes, including non-fatal and fatal injuries.

Clough et al (2015) also examined the potential for identifying systematic relativities between the value of reductions in fatal and non-fatal crashes, and between crashes and other non-market values. They suggested further research is required in this area. Additional recommendations included examining a wider set of issues affecting estimates of the value of risk reduction, including *the situation*, eg between when respondents have some control (as in car driving) and those in which they do not (as passengers in public transport), acute versus latent risk, eg risk of a sudden (crash) and delayed (air pollution) death, and *respondent characteristics*, eg age. This is a large set of issues, only some of which we have explored in this study. The research examined how a single survey could be used to identify multiple values, particularly those of fatality and injury risk reduction and of travel time.

Clough et al (2015) suggested the methodology used to update the value of reductions in fatal crashes should be a stated preference (SP) valuation study, either using the contingent valuation (CV) method or choice modelling (CM). This study investigated various methodologies, including CM which was subsequently selected as the approach for the pilot study.

This report summarises literature on survey methodologies and methods for non-market valuation in the transport sector. It discusses possible survey structures and different values to be included. It sets out how valuations can be derived based on survey results.

2 Non-market values in transport

Governments have limited (financial) resources to tackle problems affecting transport outcomes, including journey times and the risks of crashes with injuries or fatalities. Road transport investment and operational decisions weigh up the effects of project and programme options. Where options with lower risks or journey times have higher costs, the decision involves a trade-off between these factors, eg how much additional cost are we willing to bear to reduce the risk of injuries or fatalities, or how much are we willing to have a longer travel time if it will increase safety?

In this section, we explore literature that has addressed the valuation of attributes of interest for this study. The focus is on valuations made by road users, such as car drivers, because they are the ones making regular decisions which trade-off transport outcomes. Some of the nonmarket values associated with road projects and road use, particularly the environmental impacts, are borne by the wider community. These are external effects which would not be expected to be weighed up by road users in the same way as the effects they face directly. Those borne more directly include crash risks, the value of time and journey time reliability, road condition and aesthetics (scenic value of the trip).

Below we address the following non-market values in turn:

- fatal crashes
- non-fatal crashes
- value of time
- journey time reliability
- other values, including health, environment, road condition and amenity.

2.1 Fatal crashes

2.1.1 Valuation approaches

There are three main approaches to placing a value on reductions in fatalities (Mishan 1982; Viscusi 2005; Risbey et al 2007; Clough et al 2015):

1 Compensation payments (or insurance value) are an amount paid to someone after an incident that will provide them with the same level of utility (wellbeing) as before. In the case of a fatality (where the person who dies cannot be compensated), the person's insured value can be assumed to represent the estimated value.

2 Human capital approach (HCA) measures the value of a lost life (due to premature death) as the discounted stream of future earnings of that individual.

3 Willingness to pay (WTP) for incremental changes in the risk of a fatality.

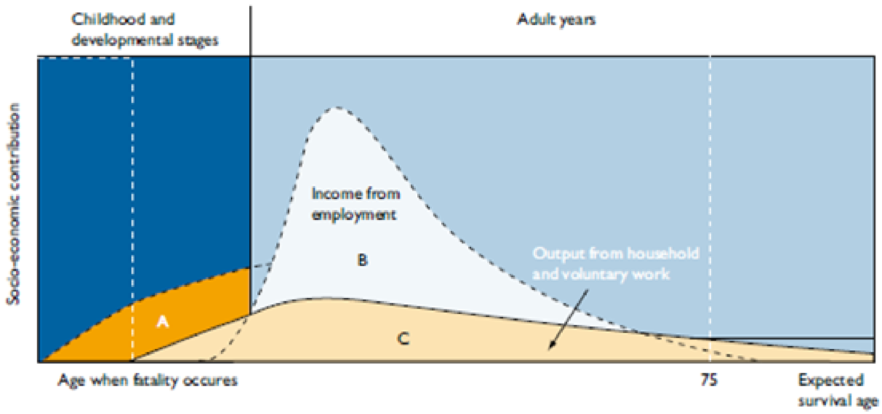
Mishan (1982) also includes (and dismisses) the politically defined approach, in which value can be derived from the government's expenditure on risk reduction as a reflection of society's valuation.[1]

[1] This is of no interest to this research because it is self-defining: the government would be using its own previous decisions to justify its future decisions.

The HCA was initially the most common approach (Mishan 1982) but changed towards WTP-based estimates of the value of risk reductions in the 1980s (Hensher et al 2009). Most economists have advocated the WTP approach since the work of Schelling (cited in Viscusi and Aldy 2003) on the economics of life saving. WTP and HCA adopt quite different philosophies.

The WTP approach is an ex-ante (from before) valuation from the perspective of the individuals whose lives are being valued. HCA is an ex-post (from after) valuation, in that it estimates the foregone output from society's perspective after the fatality occurs, which might take into account some or all of the contribution over time (figure 2.1). In other fields, we can imagine ex-ante and ex-post valuations could be similar. An ex-ante analysis would be imagining the world after the event and, given perfect foresight, it could be assumed the valuations would be the same as ex-post valuation. However, there is a fundamental difference relating to fatalities, in which the ex-post world is unimaginably different for individuals valuing their own lives.

Figure 2.1 Contributions to society across the normal lifespan of an individual



Notes: Area A = losses due to premature death of a child; B = potential losses in income from employment; C = non-wage contributions forgone by society due to premature death
Source: BITRE (2009)

WTP is generally regarded as superior because it is consistent with approaches taken to valuing other goods and services from a welfare (wellbeing) perspective, ie it measures value based on relative preferences. It is also consistent with the decision perspective being taken, which is at a time when the people whose lives are being considered are all still alive.

The focus of this research was on the WTP approach.

2.1.2 Value of statistical life

Measures taken to reduce the number of fatal crashes do not eliminate fatalities but reduce the probability or risk of a fatality. Approaches to valuation estimate the value to society of marginal changes in the risk of death, ie incremental changes in the level of risk. These values are referred to as the value of statistical life (VoSL or VSL) (Viscusi 2005) which is similar to, and sometimes interchangeable with, the term value of a prevented fatality (Thomas and Vaughan 2015).

Analysts are at pains to point out the concept is not about placing a value on the life of any individual, but about identifying the value to society of reducing the risk of fatalities. Nevertheless, it uses individuals' specifications of the value they place on their own lives. Viscusi (2005) writes:

> *the key question is how much are people willing to pay to prevent a small risk of death? ... This risk-money tradeoff provides an appropriate measure of deterrence in that it indicates the*

individual's private valuation of small changes in the risk. It thus serves as a measure of the deterrence amount for the value to the individual at risk of preventing accidents and as a reference point for the amount the government should spend to prevent small statistical risks.

The VoSL is a statistical construct that extrapolates from the value placed on reducing the risk of a fatality from one level to some lower level. For example, an intervention which reduces risk by one in a hundred thousand on average across a population of 100,000 people can be described as saving one statistical life (Robinson 2007). If individuals, on average, are willing to pay $50 to reduce risk by one in 100,000, this is equivalent to a VoSL of $50 ÷ 1/100,000 = $50 x 100,000 = $5 million. Mathematically, the VoSL is expressed as:

$$VoSL = \frac{WTP_{\Delta r}}{\Delta r} \qquad \text{(Equation 2.1)}$$

where $WTP_{\Delta r}$ is the WTP for a change in the level of risk and Δr is the change in risk level.

2.1.3 Approaches to measuring VoSL

In most countries, estimates of the VoSL are based on estimates of WTP to reduce the risks of death (Clough et al 2015), with estimates derived from studies of stated or revealed preferences (OECD 2012):

- Stated preference (SP) methods, which include contingent valuation (CV) and CM approaches, use surveys with hypothetical valuation scenarios to derive values of changes in mortality risk, from which the VSL can be derived.
- Revealed preference (RP) techniques derive the WTP by examining market responses to risk levels. Chiefly these are:
  - 'hedonic wage' (HW)/wage risk studies – differences in wage rates reflecting differences in workplace mortality risks
  - 'averting costs' studies of markets for products that reduce or eliminate risk, eg increased expenditure on safer vehicles.

OECD (2012) notes both SP and RP methods have their strengths and weaknesses, but there has been a growing emphasis on SP methods in recent years. OECD (2012) also notes it is not clear whether either approach provides systematically different (higher or lower) results from the other. The reason for using SP is partly because there is not enough RP data available (Bishop 2003), but there is some theoretical advantage also: as Bishop (2003) notes, 'true economic values are unobservable', eg they include consumer surpluses.

Until recently the CV method has been used most commonly, but use of CM is increasing (OECD 2012). SP approaches provide survey respondents with discrete choices (which differ in their mortality risk) with different costs; typically, CV studies present the same choices (sets of attributes and values) to all respondents, whereas choice experiments vary these.

It is often argued the main drawback of the SP approaches is the hypothetical nature of the research, ie market expressions of value are different from SPs, and we discuss this in more detail in chapter 3.

Below we set out the development of approaches in New Zealand, followed by a discussion of current issues in the assessment of VoSL.

2.1.4 VoSL in New Zealand

2.1.4.1 Base value

New Zealand transport agencies, which include the Ministry of Transport (MoT) and the Transport Agency (formerly Transit New Zealand), have valued lives saved on New Zealand roads in cost benefit analyses (CBAs) since the 1980s (Miller and Guria 1991). Initially, values were estimated using the HCA which estimated the present value of the wage work or housework the deceased will not perform. Using this approach, the cost of a fatality was estimated at around NZ$235,000 in 1990 (equivalent to approximately $400,000 in 2016 dollar values).[2] Miller and Guria (1991) described this value as 'embarrassingly conservative'. The key criticisms of HCA is its use of market values for wages rather than market values for safety, and its tendency to undervalue the lost joy of living, and the lives of women, children and elderly (Miller and Guria 1991; Leung 2009). For example, using a 10% discount rate as in the 1990s, the value of a two-year old would be valued at less than twice their estimated annual wage from age 20.[3] Miller and Guria are also critical that the value of the elderly was measured using their superannuation earnings.

These criticisms, and the widescale adoption of WTP methods in other countries, led to the adoption of WTP in New Zealand. A report by Quigley et al (1989, cited in Miller and Guria 1991) used the results of US studies to derive a value of NZ$800,000 per life (NZ$1.45 million in 2016$), but a workshop convened by the MoT and the National Roads Board recommended deriving New Zealand-specific values.

This led to the inclusion of a set of questions to derive WTP into the MoT Household Travel Survey (HTS) of 1989/90.

To ensure reliable VoSL estimation, questions related to reductions in risks to different people (eg the respondent, their family, or other people) were provided through different hypothetical goods/services (eg safer toll road, road safety course, car safety features, safer neighbourhood and extra taxes for road safety) (Miller and Guria 1991).

Results from the 1989/90 survey were used to recommend a VoSL of $2m, an order of magnitude greater than the HCA-based estimate. The government used this recommendation to set the VoSL for all transport sector evaluations at $2 million in 1991 dollar values (Leung 2009), with annual adjustments being made to account for changes in the average hourly earnings.

A repeated survey in 1997/98 suggested a new VOSL of $4m (NZ$1998), significantly higher than the wage inflation-adjusted value of $2.5m (Guria et al 2003). Due to 'some unresolved policy issues', the government did not adopt the revised VoSL estimate and the VoSL continues to be based on the value established in 1991 (Leung 2009). One of the arguments against adoption of the higher value was it would result in a shift of resources away from investments in travel time savings and decongestion, and toward road safety features (Clough et al 2015). However, the updated survey's estimate of the ratio of the cost of severe and minor injuries with respect to the VoSL was adopted. This was for the pain and suffering component of the value for avoiding a serious non-fatal road injury to be 10% of the VoSL, and the value of avoiding a minor injury being 0.4%.

2.1.4.2 Updating

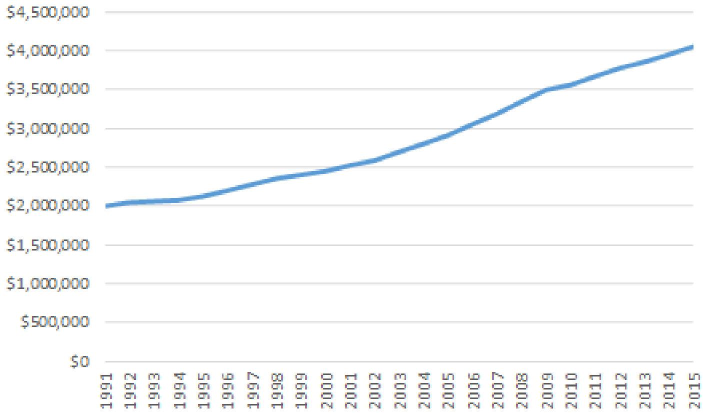
The original VoSL ($2 million in 1991$) is updated annually using average wage rates. This uses Statistics NZ's series averaged for all industry and for ordinary time. The methodology and data are explained in MoT (2016b). Currently the VoSL is calculated as $4.06 million (figure 2.2). Updating using a consumer price index (CPI) rather than wage rates would result in a 2015 value of approximately $3.26 million, approximately 20% below the wage-based updating. The value is indexed to the wage rate to keep consistency with the value of travel time, which is indexed to wages (Miller and Guria 1991). This is

[2] $396,809 if escalated from June 1990 to September 2016 using the consumer price index (Statistics NZ)

[3] For example, an annual wage rate of $40,000 (staying constant in real terms from age 20 to 65) would yield a discounted present value at age two of $78,000.

consistent with approaches used in other countries, eg in the US where studies have found VoSL is elastic with income (Trottenberg and Rivkin 2013).

Figure 2.2 Value of statistical life New Zealand 1991–2015 (nominal values)



Source: Ministry of Transport (2016b) and equivalent reports in years 2006 to 2015; for years 1992 to 2005 calculated using average wage rate (Statistics NZ Average Hourly Earnings by Industry) to update 1991 figure ($2 million)

Leung (2009) discusses other methods for updating the VoSL. One approach is to derive a New Zealand value from international values. Leung notes international research has found a strong relationship between VoSL and income (as GDP per capita). She identified formulae based on GDP per capita in research by Miller (2000) and by McMahon and Dahdah (2008), which she used to estimate the VoSL (in 2008 dollars) would have been $4.9 million and $2.96 million respectively; these values compare with the updated value for 2008 of $3.35 million. Leung developed a regression equation to estimate the VoSL for New Zealand:

$$ln(VoSL) = \alpha + \beta.\ln(GDP\ per\ capita) + \delta.\ln(DeathRate) \qquad \text{(Equation 2.2)}$$

Leung (2009) used this to develop three separate models: 1) a regression analysis using data from 10 developed countries; 2) an alternative model without the constant (α); and 3) as for 2) but without Belgium and France, regarded as atypical. Using the three models, she estimated VoSLs in 2008$s of $3.19 million, $3.42 million and $3.70 million (table 2.1).

Table 2.1 Regression results – VoSL estimates using GDP per capita

Model	α	β	δ	Adjusted R2	VoSL estimate (US$2008M)	VoSL estimate (NZ$2008M)	VoSL estimate (NZ$2009M)
1	-2.016	1.526*	0.426*	0.40	$2.0	$3.19	$3.33
2		1.335***	0.426***	0.47	$2.16	$3.42	$3.57
3		1.315***	0.548***	0.96	$2.34	$3.70	$3.87
Official value						$3.35	$3.50

Notes: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Source: Modified from Leung (2009)

2.1.5 International approaches and comparisons

OECD (2012) reviews international literature, noting the main difference among OECD countries is the reliance on RP methods (wage risk studies) in the US (where most such studies have been conducted), while Europe, Canada and Australia rely more on SP methods. Based on the international studies, the

OECD recommends a VoSL for OECD countries of US$3 million (in 2005 US$) in a range of US$1.5 million to US$4.5 million. Below we examine approaches used in Australia, the US and the UK.

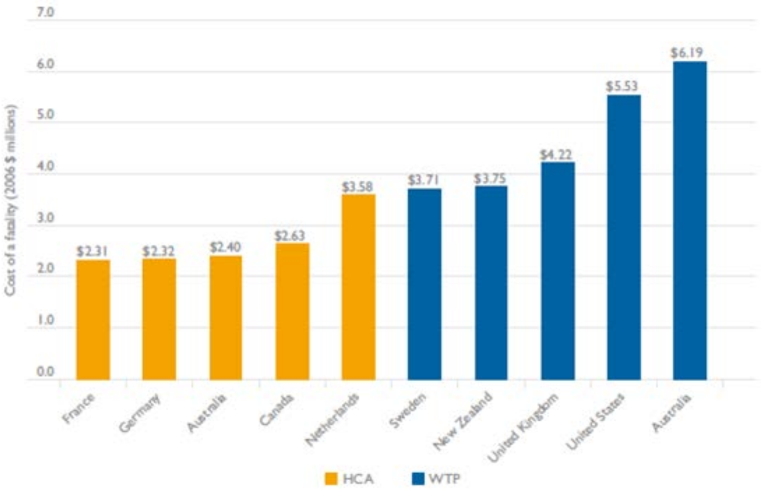
2.1.5.1 Australia

The Australian Department of Prime Minister & Cabinet (PM&C) published a best practice guidance note on the VoSL in 2014 (PM&C 2014). It suggests that ideally the VoSL would be regulation specific, taking account of the risks addressed and the people affected, but this is likely to be too costly. It notes studies that have produced results in the range of A$3 to A$15 million, but recommends a value of A$3.5 million (in 2007$ values) to be updated using CPI. The value is from Abelson (2008) and is based on a review of international studies, which included values derived from SP and RP studies.

Estimates of the cost of road crashes have been made by the Bureau of Infrastructure, Transport and Regional Economics (BITRE) using a hybrid HCA (see BITRE 2009). Building on an approach described in BTE (2000), this methodology starts with a notional value for the quality of life that would be lost by the unknown individual in the event of their premature death, adjusts it to ensure small or zero values are not adopted for children or the elderly, and adds costs for employers of work disruption, medical and hospital costs for fatally injured persons, emergency services costs and coroner investigation costs. It also adds the cost of a premature funeral, costs of prosecuting and imprisoning people for culpable driving offences, and household losses of those serving a custodial sentence. It includes an allowance for the pain, grief and suffering endured by the family and relatives of the deceased. The estimated cost of a fatal crash was A$2.67 million in 2006$. Austroads (2012) summarises the regular updating of the cost estimates on the basis of wage rates (for value of labour) and CPI for other costs (including CPI components for health and vehicle costs) (Naude et al 2015).

In contrast to the hybrid HCA, the Roads and Traffic Authority (RTA), now part of NSW Roads and Maritime Services, used a WTP study (Hensher et al 2009) to produce an estimated VSL of A$6.4 million in 2007 prices (see Naude et al 2015). The WTP values are higher than those based on HCA (figure 2.3).

Figure 2.3 Comparison of the cost of fatality estimates for developed countries (Australian $)



Source: BITRE (2009)

Tsolakis et al (2011) noted the general consensus on the adoption of the WTP as the theoretically superior approach and recommended Australia move towards using the WTP approach, including making use of values developed in New Zealand. A methodology for a WTP approach in Australia was subsequently scoped by Naude et al (2015), who also suggested using VoSL values in the interim prior to a national WTP study. The recommended values were updated versions of those developed for the RTA by Hensher et al

(2009) and as included in Austroads 2012. These were VoSLs of A$6.9 million and A$6.8 million for urban and non-urban areas respectively. Naude et al suggested these be updated using CPI or per capita GDP, multiplied by an income elasticity.

2.1.5.2 USA

The US Department of Transportation (DOT) suggests recent empirical studies indicate a VoSL of US$9.6 million (in 2016$ values), within a range of US$5.4 and US$13.4 million (Moran and Monje 2016). The values were based largely on the results of studies using data from the Bureau of Labor Statistics Census of Fatal Occupational Injuries, ie hedonic wage (wage-risk) studies. This included values from a US EPA White Paper (US EPA 2010) and a number of other studies. The guidance recommends updating these values on the basis of median real wages, ie using median usual weekly earnings in constant dollars, combined with a CPI index. However, as the constant dollar earnings numbers are estimated using CPI also (Bureau of Labor Statistics 2016), this is equivalent to simply using wage rates in current dollars, as used in New Zealand.

Moran and Monje note the usual assumption of a linear relationship between risk and WTP[4] breaks down when the annual WTP becomes a substantial portion of annual income, suggesting the assumption of a linear relationship is not appropriate for large risks. The DOT guidance notes:

> *Prevention of an expected fatality is assigned a single, nationwide value in each year, regardless of the age, income, or other distinct characteristics of the affected population, the mode of travel, or the nature of the risk. When Departmental actions have distinct impacts on infants, disabled passengers, or the elderly, no adjustment to VSL should be made, but analysts should call the attention of decision-makers to the special character of the beneficiaries* (Moran and Monje 2016, p4).

2.1.5.3 UK

Following HM Treasury (2003) guidance, road fatalities have been valued in the UK using a WTP approach (Department for Transport 2016). The cost per casualty is £1.78 million in 2015£. Values are updated annually using GDP per capita (HM Treasury 2003; McMahon 1994).

2.1.5.4 New Zealand compared with others

Leung (2009) compared New Zealand's VoSL estimate for 2008 ($3.35 million) with estimates for 12 other countries,[5] based on a currency conversion (purchasing power parity-adjusted) and GDP per capita. She found the New Zealand values were ranked ninth highest on the basis of exchange rates and 6th highest on the basis of GDP per capita.

2.1.6 Heterogeneity in estimates

Niroomand and Jenkins (2016) found clear heterogeneity for values of fatality and injury across individuals. The literature includes studies which suggest VoSL is influenced by age and income. However, a relationship with age is observed largely in HCA studies (Viscusi and Aldy 2003; Aldy and Smyth 2014), whereas the results of SP surveys are more ambiguous (Morgan and Cropper 2007; National Research Council 2008; Clough et al 2015). Relationships with income (see Moran and Monje 2016; Aldy and Smyth 2014) might be useful for benefit transfer purposes, ie adjusting values derived from international literature, but there are equity objections to using these values domestically, eg in weighting safety

[4] That is, a WTP of $1,000 to reduce risk by one in 10,000 would mean a WTP of $2,000 to reduce risk by two in 10,000

[5] Austria, Belgium, Canada, Denmark, France, Germany, Netherlands, Norway, Singapore, Sweden, UK and USA

measures towards higher income individuals. Niroomand and Jenkins (2016) suggest there are no significant relationships between the value of safety (reduced injury or fatalities) and socioeconomic variables (including age, gender, education and personal income), implying the heterogeneity was linked to some unobserved factors.

OECD (2012) summarised a range of factors that might affect estimates of VoSL and made recommendations as to their use in policy applications, as shown in table 2.2. OECD (2004, cited in OECD 2012) noted the problem of surveying children, and how they have neither the cognitive capacities nor financial resources to state RPs in surveys. They noted instead the surveys of parents, and that the results applicable to children included values, higher, lower and the same as those for adults. However, OECD (2012) recommends using a higher value for children.

Table 2.2 OECD recommendations for adjusting VoSL base values

Population characteristics	
Income	No adjustment within a country or group of countries the policy analysis is conducted for (due to equity concerns). For transfers between countries VSL should be adjusted with the difference in GDP per capita to the power of an income elasticity of VSL of 0.8, with a sensitivity analysis using 0.4 (see equation 2.1 in chapter 2.1.)
Age	No adjustment for adults due to inconclusive evidence. Adjust if regulation is targeted on reducing children´s risk. VSL for children should be a factor of 1.5 – 2.0 higher than adult VSL
Health status of population and background risk	No adjustment (due to limited evidence)
Risk characteristics	
Timing of risk (latency)	No adjustment (due to limited evidence)
Risk perception (source or cause)	No adjustment (due to inconclusive evidence). Sensitivity analysis for lower values in the environment sector than in health and traffic
Cancer or dread (morbidity prior to death)	No adjustment if the regulation targets cancer risks and/or risks that are dreaded due to morbidity prior to death. Morbidity costs prior to death should be added separately.
Magnitude of risk change	No adjustment. However, since the magnitude of the risk change clearly affects the VSL, a sensitivity analysis based on VSL calculated from a risk change similar in magnitude to the policy context should be conducted. A risk change of 1 in 10 000 annually is suggested for calculating a VSL base value
Other adjustments	
Altruism and public vs private risk	No adjustment (due to limited evidence and unresolved issues). Use 'private risk' to calculate a VSL base value. Provide illustrative adjustments in sensitivity analysis.
Discount for hypothetical bias in SP studies	No adjustment (due to limited evidence).
Correction for inflation	Adjustment based on the national CPI.
Correction for increased real income over time	Adjust VSL with the same percentage as the percentage increase in GDP per capita

Source: OECD (2012)

2.1.7 VoSL for air pollution studies

The increases in the risk of death are at the same rate across all age groups, with the exception of air pollution, where premature deaths tend to have the greatest impacts on the elderly (Kuschel et al 2012). The use of VoSLs for air pollution CBAs has been debated widely because:

- On the one hand, SP studies have been ambiguous over whether VoSL should alter with age (see above). From this perspective, air pollution that mainly affects the elderly can be treated in the same way as other risks of a fatality and the effects can be measured using VoSL.
- On the other hand, WTP results are much smaller when SP studies have asked questions more directly related to the nature of the risk, ie dying earlier but at some time in the future.

Studies that have examined the WTP for reduction in the risk of air pollution deaths have focused on estimating the value of a life year (VoLY), ie the value of extending life by one more year. VoLY can be estimated in various ways, although the simplest is to convert the VoSL into a discounted stream of annual life year values over the remaining lifetime of the subject. This is the approach adopted in the CBA of the EU CAFE programme (AEA Technology Environment 2005). The formula used is:[6]

$$VoLY = \frac{r.VoSL}{1-(1+r)^{-n}} \qquad \text{(Equation 2.3)}$$

Where: r = the discount rate

n = years over which the annuity is calculated

Using this approach and the current New Zealand VoSL of $4.06 million, a discount rate of 6% and an expected life of 40 years (the average years of life remaining for the average crash fatality), results in a VoLY of $270,000.

The first survey, to our knowledge, that asked explicitly about the valuation of a gain in life expectancy was by Swedish researchers Johannesson and Johansson (1997). They administered a telephone survey of adults between 18 and 69 years old and asked the following question:

> *The chance for a man/woman of your age to become at least 75 years old is x percent. On average, a 75-year old lives for another 10 years. Assume that if you survive to the age of 75 years you are given the possibility to undergo a medical treatment. The treatment is expected to increase your expected remaining length of life to 11 years. Would you choose to buy this treatment if it costs y and has to be paid for this year?*

The resulting VoLY is between US$700 and US$1,300 in 1995 dollars (approximately NZ$1,600 to NZ$3,000 in current dollar values).[7] Half the sample had a WTP of zero; the average of positive WTP was about US$2,700. In contrast to the other values, these values are present values for some future benefit. Dolan et al (2008) discuss studies with similarly low values, eg £242–£508/VoLY (c NZ$900 to 1,890)[8] in a 2004 UK study and a Swedish study that found a low WTP for cigarettes with lower health risks that would extend life, implying a VoLY of NZ$4,400 to $10,700 in today's dollar values.

The CBA for the UK's Air Quality Strategy notes that work by Chilton et al (2004) was the only one to have derived VoLYs directly, rather than from VoSLs. The VoLYs derived from this work ranged from £6,040 to £27,630 in 2002 prices (NZ$26,000 to $118,000 in 2016 dollars).[9] However, the analysis did not discount the future values. It started with an annual WTP[10] for a one, three or six-month extension to life. Despite being derived from surveys of adults, these values were multiplied by a life expectation of 78 years without discounting, and grossed up to 12 months (one-month value multiplied by 12). If the values are,

---

[6] This formula is represented by the PMT function in Excel.

[7] Converted to NZ$ using 1995 exchange rate and inflating using NZ CPI

[8] Conversion to NZ$ using 2004 exchange rates and inflating using NZ CPI

[9] Conversion to NZ$ using 2002 exchange rates and inflating using NZ CPI

[10] An amount a person would be willing to pay each year

instead discounted (at 4%) over 40 years (expected lifetime of an adult), this would suggest a VoLY of £1,500 to £7,000 in 2002 prices (NZ$6,500 to $30,000 in 2016 dollars).

More recently, surveys in a number of European countries have been undertaken to suggest an EU-wide VoLY of €40,000 in 2010 (NZ$77,000 in 2016), but with the value varying with income across the EU (Desaigues et al 2011). However, these values too were derived without discounting. Using the same modified approach as above suggests values using their data of NZ$2,100 to $6,400 per VoLY based on 2010 WTP of €14 to €42/month to achieve a three-month life extension.[11]

These results, combined with the inclusion of lagged benefits in which reductions in air pollution do not result in immediate reductions in fatality rates because of the level of respiratory system damage (frailty) (see discussion in Denne and Atkins 2015) suggest VoSLs are unsuitable for direct application to air pollution CBAs, consistent with advice from UK Committee on the Medical Effects of Air Pollution (COMEAP 2010).

2.2 Non-fatal crashes

The first studies evaluating the economic costs and benefits of various health and medicine interventions emerged in the 1960s, and the first major injury cost studies in the 1970s (Wren and Barrell 2010). These mainly focused on the costs of industrial injuries in the context of regulatory reform of workplace health and safety legislation in the US and the UK. The first cost of 'all injury' study was undertaken by Rice et al (cited in Wren and Barrell 2010) in the late 1980s for the US Congress in the context of debate about the need for decisions on road safety expenditure.

In New Zealand, shortly after the establishment of the Accident Compensation Corporation (ACC), the first cost of injury study was undertaken by Professor Berkowitz from Rutgers University. It focused on the cost of work-related injuries and recommended the introduction of experience rating to set ACC levies (Berkowitz 1979 in Wren and Barrell 2010).

As discussed in section 2.1.4, based on Guria et al (2003), practice in New Zealand has been to adopt ratios of 10% of the VoSL for the costs of a serious injury and 0.4% for a minor injury, comparable with international findings (Leung 2009).

In the US, the costs of non-fatal injuries are scaled relative to VoSL (Moran and Monje 2013). Each type of crash injury is rated (in terms of severity and duration) on a scale of quality-adjusted life years, compared with perfect health. The abbreviated injury scale (AIS) developed by the Association for the Advancement of Automotive Medicine (AAAM) is used to group the values (see table 2.3).

Table 2.3 Relative disutility factors by injury severity level

AIS scale	Severity	Fraction of VoSL		AIS scale	Severity	Fraction of VoSL
1	Minor	0.003		4	Severe	0.266
2	Moderate	0.047		5	Critical	0.593
3	Serious	0.105		6	Unsurvivable	1.000

Source: Moran and Monje (2016)

[11] We discount this over 40 years at 4%.

The survey work for this study examined appropriate relativities for New Zealand, using the classification of injuries set out in table 2.4.

Table 2.4 Injury and fatality definitions

Injury scale	Definition
Fatal	A death occurring as the result of injuries sustained in a road crash within 30 days of the crash. There are some exclusions: • Deaths that do not occur on a public road or road to which public has access, eg race track, farm • Deaths that did not result from injuries sustained in the crash, eg a heart attack • Deaths on the road where a motor vehicle or cyclist was not involved.
Serious	Fractures, concussion, internal injuries, crushings, severe cuts and lacerations, severe general shock necessitating medical treatment and any other injury involving removal to and retention in hospital.
Minor	Injury which is not 'serious' but requires first aid, or which causes discomfort or pain to the person injured, eg sprains and bruises.
Non-injury	Property damage only.

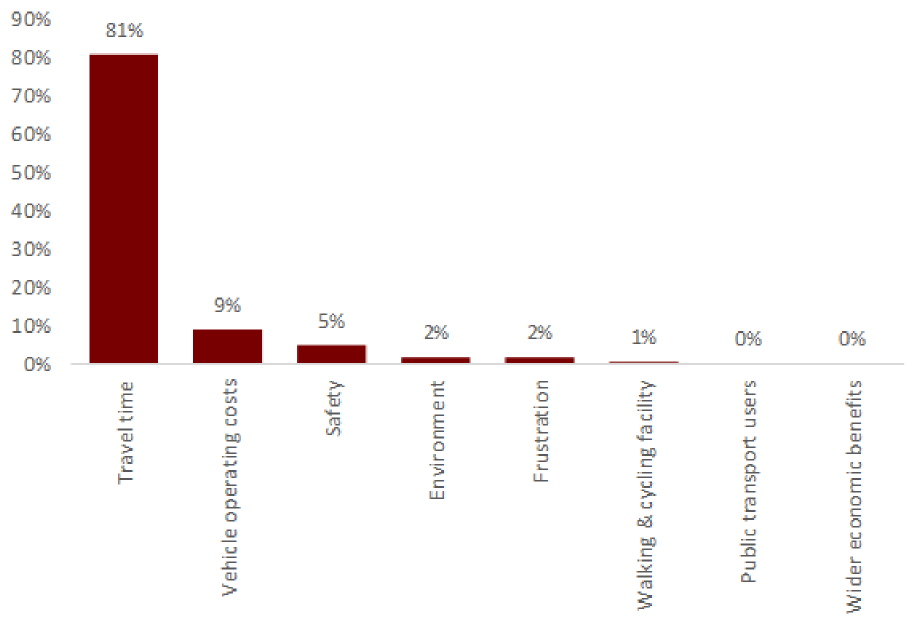
Source: Land Transport NZ (2004); MoT (2016a)

2.3 Value of time

2.3.1 Values

In the EEM, the unit values of travel time savings (VTTS) are derived from market research surveys with the results expressed in 2002$ values. Based on these values, across a wide range of New Zealand projects, travel time savings are estimated to account for over 80% of 'conventional' economic benefits of road projects (figure 2.4).

Figure 2.4 Percentage of transport project benefits per benefit stream (weighted over all projects)



Source: adapted from Wallis et al (2015)

Wallis et al (2015) provide a literature review of approaches to VTTS and we do not repeat this work here. They note the only New Zealand evidence is from SP surveys of car drivers in 2001 (BCHF et al 2002 in Wallis et al 2015).

Surveys to identify the value of travel time are not straightforward. CM, particularly SP studies, appears to be a favoured method of valuing travel time savings in much of the literature. This is because individuals tend to find assigning monetary values to travel time savings difficult (for example, using WTP). Often assigning value to travel time savings is not something individuals have considered before, so when asked to do this they are more likely to indicate a value with a low level of reasoning behind it, or simply refuse to answer (Accent and RAND Europe 2010).

Studies have found respondents may experience difficulties when considering the value of time savings presented in relatively small amounts (Metz 2008). For example, when asked how valuable a two-minute saving is in travel time per day, respondents are likely to say this time saving is not very valuable. But when asked how valuable 10 minutes a week is, they are more likely to say it is a valuable time saving. Respondents tend to write off and discount smaller or seemingly insignificant time savings without considering how they may aggregate into a more substantial savings over time. This emphasises the importance of how attributes are presented in the context of SP studies (Metz 2008).

Not surprisingly, respondents also tend to value travel time savings in the context of their overall journey duration (Metz 2008). For example, a time travel saving of five minutes may or may not be valuable depending on how long the overall journey duration is. Saving five minutes on a trip expected to take 20 minutes is likely to be considered a valuable saving, but if the trip is expected to take two hours, it is likely to be considered less valuable (Metz 2008).

2.3.2 Relativity between safety and travel time

Theory suggests a statistically significant interaction exists between the value of safety and the value of time (Meade and Cheung 2016). When road users drive faster they are trading-off travel time (and any other benefits of speed) against the risks of crashes. Evidence shows a marginal increase in travel speed results in a more than linear increase in fatal or injury crash risk (Patterson et al 2000). Given the importance of these two elements (figure 2.4), optimal resource allocation strongly depends on the weights assigned to them (Guria 1990), eg do we make roads which are faster or safer?

Clough et al (2015) examined literature on the relationship between safety and travel time and suggested it might not be stable because of the number of factors that affect the value of time. We discuss these factors below.

2.3.3 Factors affecting value of travel time

The VTTS can vary with factors including travel mode, trip purpose, time of day, time spent, income, gender, household composition, age, travel time variability, comfort and speed (O’Fallon and Wallis 2012).

Whereas the starting assumption is usually that all travel time is a cost, a New Zealand survey reported by O’Fallon and Wallis (2012) found some people expressed a preference for a greater than zero commute time, partly explained by the benefit gained from the time, eg as time to think, listen to music, enjoy the scenery and or the exercise. However, others expressed a disutility because of traffic conditions or the waste of time. They concluded that SP surveys often do not consider people may want to travel, may value their travel time, and may not choose faster transport modes just because of their speed advantage.

Wallis et al (2015) examined the relationship between the VTTS and trip duration, extent of the saving and income. From a review of literature, they found:

- A suggestion of a strong and positive relationship between unit VTTS (eg $/minute) and trip duration, ie savings are valued more highly the longer (in distance or time) or costlier the trip. Values per minute are 50–100% greater for longer (3+ hours) than for short trips (<20 minutes). However, they note studies are not uniform on this issue, and that some suggest VTTS reduces with trip distance.
- Unit VTTS increases with the size of the time saving up to a threshold level of time savings. However, they dismissed this on the basis of 'the artificial nature of the experimental design':
  - The short-term focus of SP methods. Respondents are likely to assess the options presented as of a one-off, short-term nature, so they perceive little or no opportunity to adjust their behaviour to make good use of any time savings. In the longer term, it might be expected that a regular time saving would become incorporated into people's schedules, so they could secure significant benefits from it.
  - Task simplification. When making choices in an experiment, people will tend to ignore time savings that are 'too small to matter' (eg very small compared with the overall journey time). This will be an issue particularly when survey respondents are presented with a mix of larger and smaller time savings.
  - Treatment of variability. Respondents assume the time savings will be imperceptible relative to the day-to-day variability in the time taken for typical trips.
- Unit VTTS varies with income and with trip circumstances, eg it will be higher when people in a hurry to get somewhere will self-select for faster modes (cars/taxis rather than buses).

Richardson (2003) found, all else equal, private transport users had a much higher value of time (VOT) than public transport users, and zero and positive values of travel time are more common among income-poor or time-rich travellers. Similarly, Salomon and Mokhtarian (1998) found VOT was positive for a minority of the population, but was negative on average across the population as a whole.

Hess et al (2005) doubt the credibility of the above findings. First, they highlight that most CMs are applied under the assumption errors are normally distributed around the mean VOT estimate. However, if the true VOT distribution has a mean close to zero with a long negative tail, use of a symmetrical normal distribution is inappropriate, and may bias VOT estimates. They suggest choice modellers apply a bounded distribution, where bounds are estimated from the data, as demonstrated in Niroomand and Jenkins (2016) and Hensher and Greene (2003).

Hess et al (2005) also claim evidence of VOT varying by transport mode is due to absence of some travel-experience attribute(s) from the utility specification. They theorise, if a CM (or utility function) accounts for all travel-experience attributes, 'only the actual cost in time as a resource would remain; this would be constant across alternatives (eg modes or activities) for a given person at a specific moment in time, yet would most probably vary across individuals and across the time-of-day' (p10). SP CM by Tseng and Verhoef (2008) provides empirical support for VOT varying by time of day.

Hensher and Greene (2003) analysed long-distance travel survey data from New Zealand to find VOT varies by trip length. This is consistent with the findings of Arup, ITS and Accent (2015). This type of heterogeneity differs from that described by Richardson (2003), as it implies VOT is correlated with another route attribute rather than an individual's socioeconomic status. Hensher and Greene (2003) identified this heterogeneity by interacting the travel time savings variable (set as a random parameter) with the trip length variable (set as a fixed parameter).

Assuming choice sets are described by a broad set of attributes and modelled by a mixed logit[12] model, we expect to find a negative VOT coefficient, which is likely to vary across individuals, the time of day and trip length.

2.3.4 Approaches to valuing time

Procedures for valuing time are set out in the EEM (NZ Transport Agency 2016). The approach to valuing time has changed recently (Wallis et al 2015): until 2013, time savings were derived from market research surveys and analyses undertaken by Beca Carter Hollings & Ferner in 2001 (cited in Ian Wallis Associates Ltd 2014), benchmarked against international evidence. The values were based on SP pair-wise comparisons, which asked drivers to express preferences between two routes with different travel times and costs. The research was used to derive values differentiated by mode and trip purpose, and by travel quality and comfort factors (eg congested conditions, standing vs seated public transport passengers). From 2013 a single ('equity') base value has been adopted across all modes.

2.4 Journey time reliability

In addition to travel time savings, road users value trip reliability, ie the extent to which they can have certainty over trip length (reduced variability). This applies both to car drivers and passengers, and to users of public transport and freight services.

There are different ways in which variability occurs, eg Bates et al (1987) in Li et al (2010):

1. Inter-day variability caused by seasonal and day-to-day variations (such as demand fluctuations, crashes, road construction and weather changes)
2. Inter-period variability which reflects the impact of differences in departure times and the caused changes in congestion
3. Inter-vehicle variability mainly due to individual driving styles and traffic signals.

The most common approach to measuring the value of reliability (VOR) is the mean variance model. It assumes people would prefer to leave later and arrive earlier, and utility (wellbeing) is accumulated at home and at work, but not during travel (see, for example Fosgerau 2016). This approach, which is adopted by the EEM, measures the value of improving variability as a change in the standard deviation of travel time for a route and the value of time. In the EEM (appendix A4.5) improvements in reliability are valued as 0.9 times the product of the value of travel time ($/h) and the change in the standard deviation.[13]

Li et al (2010) describe two other models:

- The scheduling model. It considers the consequences of unreliable travel time. Unlike the mean-variance model which assumes travel time variability leads to the loss of utility by itself, the scheduling model considers disutility is incurred when not arriving at the preferred arrival time (PAT), either early or late. The model uses separate parameter values for variability either side of the PAT.

---

[12] Also known as the random parameters logit model (RPL) or mixed multi-nominal logit.

[13] The 0.9 factor is the value of reliability based on a typical urban traffic mix. For projects with a significantly different vehicle mix, it is suggested evaluators use 0.8 for cars and 1.2 for commercial vehicles.

- The mean lateness model is becoming the 'standard' approach for analysing reliability for passenger rail transport in the UK. Travel unreliability is measured by the mean lateness at departure and/or arrival, while the mean earliness (negative lateness) is not considered.

Recently, empirical studies have demonstrated that travel time reliability has a significant impact on travel decision making (eg departure time, route, or mode choice). In the Netherlands, improving travel time reliability is regarded as a primary objective for the Ministry of Transport, Public Works and Water Management in the coming decade (AVV 2004). The UK government aims to achieve a 25% reduction in train delays over 30 minutes, and improve the reliability of train services from 88% in 2007 to 92.6% by 2014 (Department for Transport 2007).[14]

Li et al (2010) note some studies estimated higher values for reducing travel variability than for reducing the scheduled journey time or the average travel time. Where measured, studies have shown travel time reliability to be valued anywhere from half the value attributed to time savings[15] to double that of the VTTS.[16]

Small et al (1999) found travellers with children have a higher disutility associated with lateness, and those with lower incomes have less disutility from early arrivals. Tilahun and Levinson (2010) found higher income individuals valued reliability less; they reasoned this was because of the greater flexibility provided by higher income jobs. Small et al (1999) also found the 'value of being early' variable to be statistically significant as a quadratic term (ie $early^2$), which implied the positive utility of being early exists up until a time of about three minutes.

2.5 Other values

2.5.1 Health

In addition to the effects of crashes, health (and psycho-social)[17] effects of transport include those resulting from (Economic Commission for Europe 2006):

- emissions and impacts on air quality
- changes in levels of physical activity
- climate change
- noise.

Except where they represent changes in levels of physical activity, the health effects listed are externalities of the transport system. They affect people other than the road users themselves and are therefore unlikely to be weighed up by road users in their road use decisions. In this project our concern is with

[14] Rail reliability is measured by the public performance measure. This is not met if a scheduled train service is cancelled or arrives at its final destination more than 5 minutes late (or 10 minutes for inter-urban services).

[15] For example, Bhat and Sardesai (2006) found a VTTS of US$12.2/h compared to a value of travel time reliability of US$6.1/h, while Hensher (2001b) found values for long-distance car travel for time and reliability of NZ$8.70/h and NZ$5/h respectively.

[16] Small et al 1999 found a VTTS of US$6.30/h which compared to a value of reliability of US$12.60/h, while Batley and Ibáñez 2009, examining rail patronage, found VTTS and value of reliability to be £15.40/h and £31.80/h respectively.

[17] These are effects caused by environmental and/or biological factors on individual's social and/or psychological health.

exploring the values road users place on the benefits of different outcomes of road use, and the trade-offs between them. Given this we explore only the values relating to physical activity.

The EEM includes values for the health benefits of active modes (walking and cycling) expressed as $/kilometre walked or cycled. These are used to place a value on modal shift, eg because of new facilities. These health values are included in the overall demand for these transport modes.

Rather than using a SP survey to estimate the health benefits of active transport modes directly, Genter et al (2008) estimate the health benefits of active modes in New Zealand in physical terms (eg reductions in mortality, cancer rates, type 2 diabetes) and then use published estimates of the value of reductions in these health effects. The combined numbers are used to produce values per kilometre of walking and cycling after weighting the numbers according to the existing activity levels of the general population (table 2.5).

Table 2.5 Estimates of health benefits of active modes ($/km)

	Walking	Cycling
Low	$3.53	$1.77
Medium	$4.27	$2.15
High	$5.01	$2.51

Source: Genter et al (2008)

This is the same approach as used by the WHO in its health economic assessment tools for walking and cycling (Kahlmeier et al 2014), although it bases its assessment of benefits only on mortality-reduction benefits rather than a wider set of health effects.

2.5.2 Value of environmental impacts

As with health effects, environmental impacts tend to be external costs affecting the wider community rather than being borne by the road users themselves. This includes the aesthetic and ecological impacts of road construction plus the road run-off, noise and local air quality impacts of road use. Despite the effects being external, some road users opt for transport options that limit the environmental costs borne by the wider community and some would be willing to pay to cover the environmental costs. In this respect, people will sometimes make decisions from a community rather than an individual perspective (Sagoff 1988). However, because of the close correlation between environmental effects and other factors, it may be difficult to separate them out. For example, CO_2 emissions and fuel consumption (and costs) are exactly correlated.[18]

Daniels and Hensher (2000) use a choice experiment to estimate environmental values (including traffic noise, traffic on local street, and bushland and open space lost) alongside traditional values such as travel time and cost. In general, environmental impacts were statistically insignificant. However, a significant

[18] CO_2 emission costs are internalised in fuel prices via the emissions trading scheme. This is an efficient internalisation if we assume the price of emission units reflects the social costs to New Zealand. This is broadly correct if we assume the social costs of CO_2 are equal to the lowest costs to New Zealand of coming into compliance with its international obligations, rather than based on estimates of marginal global damage costs, or some longer run (dynamically) efficient price which might incentivise investments consistent with a higher expected future price. CO_2 costs will be fully internalised by 1 January 2019. Currently (2018) the obligation for producers or importers of liquid fuels is to surrender 0.83 emission units per tonne of CO_2-equivalent emissions; the obligation will increase to 1.0 from 1 January 2019.

relationship was found when an environmental attribute is interacted with an attitudinal dummy variable.[19] Daniels and Hensher (2000) explain that environmental attributes have a relatively indirect impact on transport users' self-interests. Therefore, such attributes are unlikely to be appropriately traded off with those 'close in self-interest proximity', such as travel time savings and travel cost.

2.5.3 Road condition

Transport route preference may be affected by road condition and the consequent impacts on in-vehicle road noise and travel smoothness. Related road condition issues will include how winding (ie not straight) the road is.

The EEM places a value on reductions in roughness, made up of reductions in vehicle operating costs and occupants' WTP to avoid rough road conditions. However, roughness might also be related to crash risks (Cenek and Jamieson 2012) for which reductions are also valued.

Hartmann and Ling (2016) studied the perceived value of road maintenance in Singapore. A survey was used which included three road condition indicators: road evenness, water ponding and road cleanliness. Respondents evaluated network performance on a 10-point scale rather than using WTP. Road evenness was the most significant factor. They also measured values for clarity of road signs and the efficiency of traffic redirection. Olsson (2002) used stated choice survey techniques to identify car users' monetary valuations of ride quality resulting from road maintenance. She found the worst pavement damage was roughness, with motorists willing to pay 1.7 SEK (Swedish crowns) per kilometre (c NZ$0.32/km)[20] to avoid roads that were damaged.

Hensher and Sullivan (2003) analyse New Zealand drivers' SPs for regional and inter-urban routes. Results show straightening out of existing roads will increase traveller utility in addition to the benefits of travel time savings.

2.5.4 Amenity

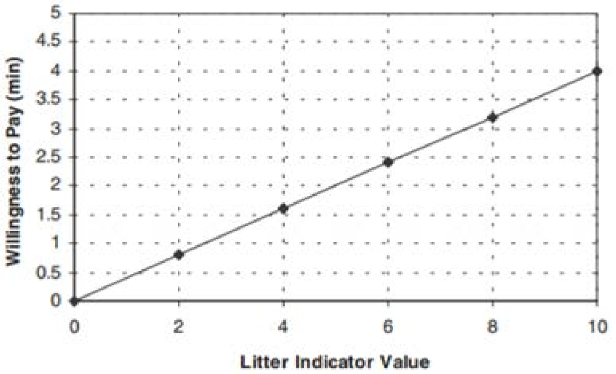
The EEM notes, where there is an attractive view from a road, the aesthetic appearance of the road to users should be taken into account in project appraisal. However, no valuation methods are discussed. It is likely that, some road users at least, will make route choices which are more scenic, even where there are higher costs (Wu and Fleming 2012).

Hyman (2004) examined the relationship between levels of litter (litter indicator value) and the amount focus group participants were willing to pay in travel time to obtain a particular level of litter removal (figure 2.5).

[19] A 'yes' or 'no' answer to the following question constituted the dummy variable: should governments do more to protect the environment, even if it leads to higher taxes?

[20] Inflated from mid-2002 to February 2017 using Swedish CPI (source = Statistics Sweden) and converted to NZ$ using current exchange rate (www.oamda.com)

Figure 2.5 Relationship between litter indicator value and willingness to pay (in minutes)



Source: Hyman (2004)

Groothuis et al (2007) examined the value of scenic views as a WTP to remove billboards that detract from views. Alivand and Hochmair (2015) and Alivand et al (2015) explored the use of CM to examine the value of scenic routes, including those that featured water bodies, mountains or forests.

We explore CM in more detail in the next section.

2.6 Factors for inclusion in a survey

In this chapter we have discussed several factors associated with road travel that affect the overall value of the trip. Of interest to this study are those relating to fatalities and injuries and to travel time and reliability. However, there is an interest in including other factors within a survey. The discussion above suggests those relating to road condition and/or the scenic value of the trip might be usefully included, but factors such as environmental effects are less relevant because these are more likely to affect those other than the route decision maker, ie they are external effects on the surrounding community.

In the next chapter we discuss CM as an approach to obtaining values for several attributes from a single survey.

3 Choice modelling

3.1 Choice modelling vs contingent valuation

CM and CV are SP methods commonly used to infer non-market values. SP techniques estimate an individual's relative preferences for individual goods and/or attributes through surveys, with the value calculated either directly or indirectly for CV methods, or indirectly when using CM. In contrast, RP techniques use real market decisions to infer values of a good or service (Access Economics 2008). Although RP methods are often favoured because of their grounding in real world transactions, their applicability to non-market valuation is limited to cases where a market exists (Haab et al 2013), eg noise values can be derived from a regression analysis on prices of houses with different exposure to road or other noise.

CV has been the traditional SP tool for non-market valuation. Early CV methods first defined a hypothetical status quo and then asked respondents directly their WTP for, or willingness to accept (WTA), a specific change to the described scenario. Later CV approaches used a theoretically incentive-compatible referendum approach to infer values from choices between two alternative states of the world, including an analyst stated difference in costs to the individual. CV has been subject to a number of criticisms, however, of which Hausman (2012) claims the most serious are:

1 Hypothetical bias – the inconsequential nature of the CV questions means respondents may not fully consider their income constraints, and thus overstate their WTP for a certain good. For example, CV methods applied in transport are typically concerned with the trade-off between money and risk, and ignore other key attributes such as travel time (The Hensher Group Pty 2007).

2 Differences between WTP and WTA – compensation paid (WTP) is generally less than compensation demanded (WTA) invalidates the CV method.

3 The 'scope' effect – people's WTP for the same good can vary depending on whether the good is valued on its own or as part of a bundle of goods.

These issues led Hausman (2012) to conclude that unquantified benefits or costs are preferred to those calculated via CV. Haab et al (2013) reference an extensive volume of literature to contest each of Hausman's 'long-standing' issues with CV methods. For example, they refer to studies that found differences between stated intentions and actual behaviour (ie hypothetical bias) can be explained by the characteristics of the study, discrepancies between WTA and WTP could be a result of people valuing losses more than gains, and they question the validity of tests used to measure the 'scope' effect.

House of Lords (2006) notes some criticisms of WTP approaches applied to the value of safety in evidence provided by John Broome to the Select Committee on Economic Affairs. According to Broome, the approach suffers from two fundamental flaws in its application to valuing safety:

1 Violation of expected utility theory – this theory states that expected utility is a linear function of probability, therefore the value of reducing a person's risk is constant. However, according to the WTP method, the value of reducing a person's risk increases with the level of risk he or she bears.

2 The apparent priority given to the safety of the wealthy – the WTP approach implies money has the same value to everybody which in effect overvalues the life of a rich person and undervalues a poor person (a criticism that might apply to CBA more broadly).

The development of CM methods for non-market valuation has largely been in response to criticism of the CV method.[21]

CM techniques quantify the trade-offs people make between specific attributes of a good or service. This is done by presenting a series of hypothetical scenarios in which respondents select their preferred choice among a number of options, otherwise known as a choice set. Options may be policy outcomes (eg different upgrades to a commuter route) characterised by attributes (eg travel time, travel speed, level of safety and level of congestion) of varying qualitative or numerical levels (Scarpa and Rose 2008). Reliable choice surveys or choice experiments simulate real-world scenarios to the extent possible and include all key attributes in the description of the good or service. For non-market valuation, a monetised attribute is included so dollar values of non-market goods can be estimated from the trade-off between, eg safety and cost.

Although some of the CV criticisms (eg their hypothetical nature) also apply to CM, choice experiments provide more information on consumer preferences than is possible through CV. For example, CM requires respondents to choose between substitute goods to more accurately reflect decision making in a real market and minimise potential for hypothetical bias and the 'scope' effect. Furthermore, unlike CV, CM can be used to estimate a good's total value and the value of attributes bundled within it (Lee 2012).

3.2 Transport applications of choice modelling

Transport applications of CM found in the literature typically relate to the prediction of transport demand, such as which route a driver will take to work. Table 3.1 summarises a number of national studies that estimated VTTS and VOR, and their sampling strategies.

Table 3.2 summaries the sampling strategies of several WTP studies focused on valuing risk reduction of road crashes. Some country-specific examples are discussed below these tables.

3.2.1 Australia

CV approaches have been the most common approach to WTP-based estimates of VoSL, but Hensher et al (2009) note CV has been criticised by human behaviour specialists and economists. They note the simulated contexts which require trade-offs between risk and money may not resemble actual trip choices where individuals have to consider a bundle of attributes, including travel time, tolls and safety with each route. In contrast, SP surveys invite participants to choose options with several attributes, which Hensher et al suggest allows the analyst to mimic actual choices with a high degree of realism. The contrast between the two approaches can be seen from two examples below (box 3.1 and figure 3.1).

Box 3.1 shows a question from the original survey used in New Zealand to establish the VoSL. It is a CV approach and asks a very simple and direct question: the respondent's WTP as a toll for a road which is safer than a free (un-tolled) road.

---

[21] CM precedes CV by two decades. CV was first proposed in 1947 by Ciriacy-Wantrup but only implemented in 1963; CM, however, was first conceived and implemented in 1931 by Thurstone (Hatton MacDonald et al 2016).

Table 3.1 National study on car VTTS and VOR: a review of sampling strategies

Nation (author)	Segment	Target sample size	Recruitment method/survey method	Sampling method
Netherlands (Significance et al 2012)	1) Commuting 2) Business 3) Other 4) Time of day	1,250 500 750 50% peak/50% off-peak	Online panel (deliver low VTTS) Intercept at petrol station, parking garage Web-based survey	Within each segment, the samples were drawn to be representative of the population (income, age, gender, region, work status)
UK (Department for Transport 2015)	1) Commuting 2) Non-work 3) Employees' business 4) Employers' business	1,000 1,000 1,000 133 (quota by firm size)	Target 80% intercept and 20% telephone; obtain 10–16% CATI samples Web-based survey	Sampling locations were selected to reasonably cover the nation and trip purposes, trip type (urban/rural/inter-urban). Intercept approach can target specific samples that were difficult to recruit (business-travel, long commute)
Denmark (Fosgerau et al 2007)	1) Commuting 2) Education 3) Leisure 4) Maintenance 5) Driver vs passenger	2,669 car respondents among 6,106 interviews; 16,791 drivers' SP and 3,837 passengers' SP observations	Online panel Web-based survey	Random via online panel supplier, supplemented by face-to-face CAPI where online panel is not well represented
Switzerland (Axhausen et al 2007)	1) Commuting 2) Shopping 3) Business 4) Leisure	~700 car respondents 6–9 SP tasks each	Recruitment method not mentioned Mail-in/mail out	Quota sampling was used to obtain enough samples for rarer segment such as long distance and business travel
Norway (Ramjerdy et al 1997)	1) Commuting 2) Other non-work 3) Business 4) Trip length	300 interviews per long distance band (30–100/ 100–300/300+ km), and 1,545 interviews for car trips shorter than 30 km	Telephone recruitment, home interview	Sampling locations were selected to cover the country Stratified sampling technique was used with samples proportional to population

Table 3.2 National WTP study on value of risk reduction for road crashes: a review of sampling strategies

Nation (author)	Segment	Target sample size	Recruitment method/survey method	Sampling method
New Zealand (Miller and Guria 1991)	1) Main urban areas 2) Remainder of New Zealand	629 respondents age 18+	Part of the HTS Face-to-face interview	Geographically stratified and random sampled at four different stages (main urban areas, meshblocks, households, person) Randomly sampling one in four households participating in the HTS

Nation (author)	Segment	Target sample size	Recruitment method/survey method	Sampling method
New Zealand (Guria et al 2003)	1) Urban 2) Rural	1,051 respondents aged 17+	Part of HTS *Face-to-face interview*	Stratified the nation into 14 local government regions with samples for each region proportional to its population, except for large (under-sampling) and small regions (over sampling), then sampled in three stages (meshblock, household, person)
Norway (Veisten et al 2013)	N/A	2,342 car users	Online panel *Web-based survey*	Geographically stratified by city size and samples drawn from all regions to be nationally representative in terms of age, gender, household size and income
UK (Jones-Lee and Spackman 2013)	N/A	Various studies, all relatively small in size (167–1,103 interviews)	*Face-to-face interview*	Samples were drawn to represent the population in terms of age, gender, income and household size
Singapore (Hess et al 2016)	N/A	5,000 households	Sampling frame purchased *Face-to-face interview*	Random sampling as 5,000 respondents is a large sample size for an island of 5.6 million people
NSW, Australia (Hensher et al 2009)	1) Urban 2) Rural	142 urban 71 non-urban	Initial contacts to recruit *Face-to-face interview*	Interview sites were selected to cover urban and non-urban trips. Quotas sampling was used to obtain samples by trip length, age, and gender
Malaysia (Yusoff et al 2013)	Four ethnicities (correlation with income)	3,000 vehicle owners across all states (50% car + 50% motorcycle owners)	Mail not successful. Approaches made to respondents in offices and public areas *Face-to-face interview*	Stratified by ethnicity and random sampling Quotas were used to obtain samples by age, gender and car use

Box 3.1 CV interview question

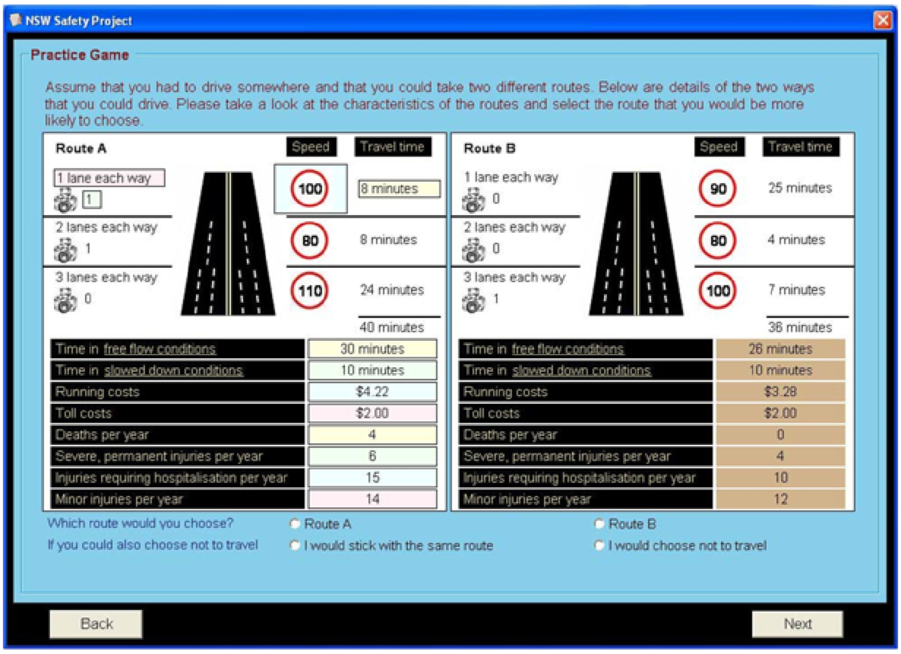
7. Imagine that you have to travel in a car for a distance of 20 kilometres each weekday for some reason. You can use two different routes – one a high risk road and the other a low risk road.

But before you can travel on the low risk road you must pay a fee – a toll. The time taken to travel each road is the same.

The toll road will reduce your risk of dying in an accident (for each year you travel) from 6 in 10,000 to 3 in 10,000. How much would you pay per one way trip to use the toll road?

Source: Miller and Guria (1991)

Figure 3.1 CM interview question



Source: Hensher et al (2009)

In contrast, figure 3.1 shows the question presented to respondents in a survey undertaken in Australia. It includes a set of different attributes for the route alternatives. Respondents are asked to choose between two routes, one of which allows faster travel (it has higher speed limits), but has a longer total journey time, higher fuel costs, and a greater risk of death or injury. Different choice scenarios systematically vary the levels of each of the route attributes to enable statistical identification of the value of each attribute.

Hensher et al used this survey with a sample size of 213 for trips in the Sydney-Bathurst region. The results were expressed as a WTP value per trip (table 3.3). They suggest the higher mean WTP estimates for non-urban settings might be because of higher speeds and the high community appreciation of the risks of crashes on the open road.

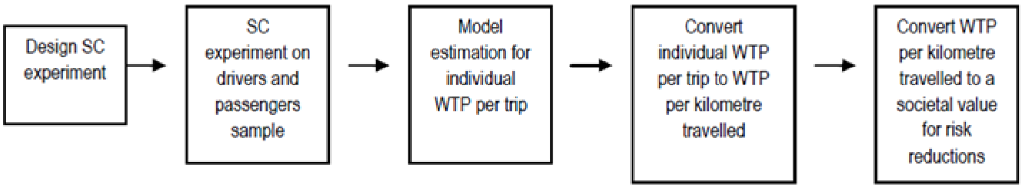
Table 3.3 WTP per trip

Attribute	Urban average (standard deviation)	Non-urban average (standard deviation)
Deaths	$0.92 ($0.31)	$3.99 ($1.12)
Permanent injuries	$0.18 ($0.05)	$0.42 ($0.06)
Major injuries (hospital)	$0.13 ($0.04)	$0.29 ($0.04)
Minor injuries	$0.12 ($0.05)	$0.25 ($0.03)

Source: Hensher et al (2009)

To convert WTP per trip values to a VoSL or a value of reducing risk of an accident, these values had to be converted into a value per person per kilometre, which were then multiplied by the inverse of the chance of death or injury per kilometre, ie the numbers of fatalities and injuries in each injury class, divided by total vehicle kilometres travelled (VKT).

Figure 3.2 Approach used in NSW study



Source: Austroads (2012)

The estimated values of risk reduction (see table 3.4) were compared with analyses of VoSL by Access Economics (2008) which average $5.0–$7.1 million.

Table 3.4 Value of risk reduction ($2007)

Attribute	Urban	Non-urban
Fatality	$6,369,655	$6,298,062
Permanent (serious) injuries	$310,292	$193,883
Major injuries (hospital)	$75,476	$56,937
Minor injuries	$16,552	$20,312
All injuries	$44,783	$48,927

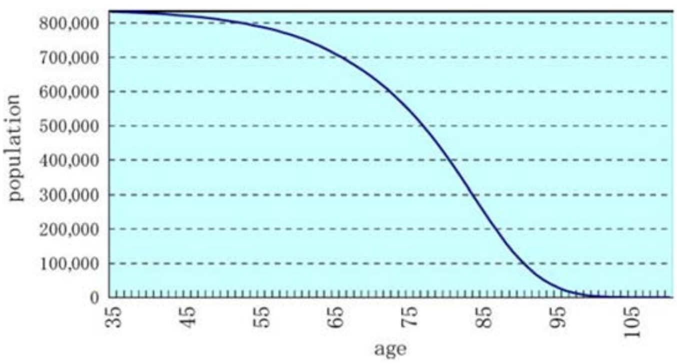
Source: Hensher et al (2009)

3.2.2 Japan

Tsuge et al (2005) used a choice experiment to estimate the marginal WTP for reducing the risk of death and used it to infer the VoSL. The purpose of this study was to understand how different risk types (eg risk of death by a crash over death by disease) and personal characteristics (eg age, income and health status) influence a person's WTP for risk reduction.

Individuals taking part in the study were first explained the notion of risk and the annual mortality rates for the major causes of death in Japan. Respondents were then shown a graph illustrating the relationship between age and survival (figure 3.3). A 10,000 cell grid illustrating risk, eg 13 coloured cells indicate a 0.13% chance of death, was then presented to further respondents' comprehension of risk.

Figure 3.3 Example of a survival curve for males 35 years and older



Source: Tsuge et al (2005)

After the above explanation was given, respondents were asked to choose their preferred option within a series of choice sets (see table 3.5). Risk reduction options are described in terms of the level of risk reduction, price, type of risk avoided, and when the good came into effect.

Table 3.5 Example choice set from Tsuge et al (2005)

<table>
<tr><th></th><th>1</th><th>2</th><th>3</th></tr>
<tr><td>Price (for 10 years)</td><td>80,000 yen</td><td>850,000 yen</td><td rowspan="4">I do not wish to purchase either</td></tr>
<tr><td>Risk reduction (for 10 years)</td><td>1/10,000</td><td>10/10,000</td></tr>
<tr><td>Risk type</td><td>Accident</td><td>Disease (cancer)</td></tr>
<tr><td>Effect starts</td><td>5 years</td><td>10 years</td></tr>
</table>

Source: Tsuge et al (2005)

Tsuge et al's analysis of the choice experiment data found individuals strongly prefer current risk reduction over future risk reduction, with an estimated annual discount rate above 20%. Their study estimated a VoSL of US$2.9 million (2005 values) for all risk types.

3.2.3 UK

Department for Transport (2015) used a CM approach to value travel time by trip purpose (eg non-work and business) and mode (eg car, train, bus). Choices within the survey were based on travel behaviour data from the national travel survey, railway passenger ticket sales and RP surveys of choices between rail trips to London.

Survey respondents (9,023 in total) were presented with hypothetical choices between a slower/cheaper travel option and a faster/dearer travel option. Controlling for the effects of travel distance and income, analysis of the choice data showed non-work travel time was valued significantly less than commuting travel time (table 3.6).

Table 3.6 Values of travel time savings

<table>
<tr><th rowspan="2">VTTS</th><th rowspan="2">Distance</th><th>Commute</th><th>Other non-work</th><th colspan="5">Business</th></tr>
<tr><th>All modes</th><th>All modes</th><th>All</th><th>Car</th><th>Bus</th><th>Other public transport</th><th>Rail</th></tr>
<tr><td rowspan="4">All modes</td><td>All</td><td rowspan="4">11.21</td><td rowspan="4">5.12</td><td>18.23</td><td>16.74</td><td>-</td><td rowspan="4">8.33</td><td>27.61</td></tr>
<tr><td><20 miles</td><td>8.31</td><td>8.21</td><td>-</td><td>10.11</td></tr>
<tr><td>20-100 miles</td><td>16.05</td><td>15.85</td><td>-</td><td rowspan="2">28.99</td></tr>
<tr><td>>=100 miles</td><td>28.62</td><td>25.74</td><td>-</td></tr>
</table>

Source: Department for Transport (2015)

The survey design allowed for the estimation of other non-market values (eg travel reliability, travel in free-flow traffic, and value of arriving early) as a multiplier of the average travel time for a given mode and journey purpose (table 3.7). For example, in the case of public transport reliability, each minute of lateness is valued at 'X' times the value of each minute of scheduled travel time.

Table 3.7 Value of travel time multipliers

Trip mode	Multiplier type	Commute	Employees' business	Other non-work
Car	Reliability ratio	0.33	0.42	0.35
	Free-flow	0.51	0.42	0.47
	Light congestion	0.72	0.68	0.83
	Heavy congestion	1.37	1.26	1.89
Bus	Value of early	-2.69	-	-3.20

Source: Department for Transport (2015)

3.2.4 Chile

Rizzi and de Dios Ortúzar (2003) used a choice experiment to estimate the VoSL for Chilean interurban motorways. Study participants were given a route choice SP survey, where each choice set contained two route options described in terms of travel time, toll charge and level of risk (annual crash rate).[22] Hypothetical choices related to a motorway journey between the cities of Santiago and Valparaiso. In every case, the trip was assumed to be unavoidable, so there was no room for a non-purchase option. Information relating to respondents' driving attitudes, experience and socio-economic characteristics were also recorded and controlled for in the CM.

The authors paid special attention to *lexicographic behaviour*, and its potential to obscure reliable estimates in CM. Lexicographic behaviour is where respondents prefer any amount of one attribute to any amount of another attribute. Of the 342 individuals studied, 150 had lexicographic preferences. Rizzi and de Dios Ortúzar (2003) give two explanations for this observation:

1 An insufficient range of attribute levels within the choice sets, eg if the cost variable was not high enough, individuals who valued risk highly would always choose the option with the lower risk of a crash, or

2 Individuals found the choices too difficult or complex, and thus simplified the decision-making process by selecting the best option in terms of one attribute.

Analysis of the choice data found a different VoSL depending on whether or not the lexicographic answers were included in the model. VoSL ranges including and excluding lexicographic answers are US$285,000 – $570,000 and US$149,000 – $206,000, respectively (2003 values). Rizzi and de Dios Ortúzar note these values should be taken with caution as their sample is not strictly random.

3.2.5 Cyprus

Niroomand and Jenkins (2016) estimated VoSL and injury from a SP choice experiment using a three-part survey. They first obtained information about respondents' current transport usage (eg mode use, frequency of using major routes, travel times, travel costs, trip purpose, typical car passengers), their current exposure to risk given the previous information, and their perceptions of transport safety risk and policy. In the choice experiment, respondents selected their preferred option from eight different choice sets, each containing two route options and a status quo alternative (table 3.8). Data from part one was

[22] Risk was described in terms of fatal crashes per year as the authors felt this is easier to comprehend than probabilities of death.

used to define the levels of the status quo option to make the choice decision more realistic. Socioeconomic questions were asked to provide control variables for the CM.

Table 3.8 Attribute levels in choice experiment

	Route A	Route B	Current route
Speed camera (per lane)	1	2	Neither route A nor route B: I prefer to stay with my normal route
Average speed (km/h)	90	80	
Travel time (minutes)	60 minutes or less	61 – 120 minutes	
Running costs (TL)	20%	10%	
Fatalities (per year)	Fewer than 10 people	10 people of fewer	
Injuries (per year)	20 people or more	Fewer than 20 people	

Source: Niroomand and Jenkins (2016)

Consistent with Rizzi and de Dios Ortúzar (2003), responses were analysed for lexicographic preferences. As such behaviour cannot be reconciled with conventional discrete CMs, all lexicographic respondents (121 in total) were removed leaving 374 useable respondents. Analysis of the choice experiment data revealed a VoSL of €717,000, and a value of injury of €16,885. Although this VoSL sits within the lower range of VoSL estimates, it better aligns with international values once national differences in income are accounted for.

3.2.6 Conclusions

CM studies have been used to define VoSLs by using choice tasks which compare different routes between origins and destinations. These generally trade-off safety (risk of injury or fatality) with speed (journey time and time reliability) and include other factors deemed important to decisions, particularly congestion. This study will use the same approach.

3.3 Choice modelling theory

In this section we set out how CM uses the survey results to produce values for VoSL and other attributes.

3.3.1 Stated choice experiments

Stated choice (SC) experiments involve *surveys* in which respondents are asked to choose their preferred alternative in each of a series of hypothetical choice tasks which systematically vary attribute levels of choice alternatives. An example is given in figure 3.1 above. Data collected from the choices is compiled and statistical analysis is used to analyse relative preferences amongst attributes.

Originating in the field of psychology in the 1930s (Thurstone 1931), SC methods have become widely used to model consumer preferences. Early proponents of SC methods made use of crude experimental designs to construct surveys in which respondents were asked to make pairwise comparisons between competing hypothetical alternatives. Since then, while the survey questions shown to respondents have changed little, many advancements have been made in the design and analysis of SC experiments. Paralleling improvements in computing and advances in econometric modelling have also occurred, allowing researchers to:

- deal with multiple choices (McFadden 1974)

- model complex forms of heterogeneity, eg the mixed multinomial logit model (Train et al 1987) and the latent class model (eg Kamakura and Russell 1989)
- model other effects associated with SC data, such as the possibility multiple choice observations (repeated measures) can be obtained for each respondent (see eg Revelt and Train 1998).

Additionally, experimental design theory specifically for SC data has advanced considerably over the past three decades (see Rose and Bliemer 2009 for a review of the literature on SC experimental design theory).

SC experiments have been shown to be capable of replicating decisions made in real markets (see eg Burke et al 1992; Carson et al 1994). In several studies (though not always), SC experiments have reproduced the behavioural outputs, such as WTP measures, obtained from RP studies (eg Cameron et al 2002; Carlsson and Martinsson 2001; Lusk and Schroeder 2004; Mark and Swait 2004; Telser and Zweifel 2002, Telser and Zweifel 2007).

Rose and Hensher (2006), Lanscar and Louviere (2008) and Hess and Rose (2009) argue the degree of realism imposed in SC surveys is one reason for this strong predictive ability. Rose and Hensher (2006) suggest the realism of SP experiments is bolstered by the respondents being asked to make a choice between a finite set of alternatives, as they would in real life. Moderating the realism is the extent to which the alternatives, attributes (eg cost, travel time, time variability) and attribute levels align with the respondent's experiences, or generally appear credible. However, for the analyst, the decision about what and how many alternatives, attributes and attribute levels to include in the SC task is often a challenging one. The decision may be influenced by what attributes and alternatives the analyst believes will systematically alter choice. Yet the literature cautions against the inclusion of too much information. Research into what constitutes appropriate choice task dimensions has tended to centre on identifying sources of cognitive burden placed upon respondents undertaking SC tasks, (eg Arentze et al 2003; Caussade et al 2005; DeShazo and Fermo 2002) as well as reducing the cognitive load placed on those same respondents (eg Louviere and Timmermans 1990; Wang et al 2001). The amount of information in a choice task will vary from one type of choice to the next, but also from one individual to the next, as tastes and motivation levels will vary. A balance therefore exists between:

- the amount of information analysts wish to gather in terms of the attributes and alternatives influencing choice
- the ability of respondents to assimilate and process information to make rational choices.

3.4 Discrete choice models

CM theory begins with the common microeconomic assumption that consumers aim to derive the greatest possible satisfaction (utility) from their limited income or budget constraint (ie utility maximisation). Accordingly, a respondent's selection within a choice set represents what maximises their utility. Using data on the alternative(s) chosen in a choice task, discrete choice models (DCMs) can be estimated. This involves the estimation of a series of regression-like equations which predict the utility the decision maker assigns to each of the alternatives. The utility is *latent* in the sense it is concealed to the analyst who must model it indirectly based on a number of observables. The modelled utilities for the alternatives are then compared to produce a probability of each alternative being chosen, based on assumptions about the random component of utility. DCMs involve the simultaneous estimation of several equations, up to the number of alternatives in the data. Unlike linear regression models, however, these equations do not directly predict the observed outcome, which in this case would be the observed choices. Rather, the 'regression-like' equations predict the latent utilities for each of the alternatives, which are then subsequently used to predict the choice outcomes.

CM draws on two other consumer behaviour theories to model decision making: the Lancasterian consumer theory and random utility theory (Lee 2012).

- Lancasterian theory suggests consumers derive utility from the attributes of a good as opposed to the good as a whole. Hence, alternatives within a choice set are described in terms of their key attributes.
- Random utility theory proposes an individual's utility can be divided into an observable or measurable component (V_j) and a random, unobservable component (e_j). Assuming these components are additive, we can define utility (U_j) derived from a specific good (j) as:

$$U_j = V_j + e_j \qquad \text{(Equation 3.1)}$$

The random element (e_j) of utility is included because evaluators cannot perfectly predict a person's utility, as it is unlikely we can observe or measure every characteristic of the individual, good, or situation that affects decision behaviour. The measurable element of utility (V_j) for good j is commonly referred to as the deterministic component, and can be explained by the attributes of any given alternative (ie good):

$$V_j = \beta_{0j} + \beta_{1j}.X_{1j} + \beta_{2j}.X_{2j} + \ldots + \beta_{kj}.X_{kj} \qquad \text{(Equation 3.2)}$$

where:

V_j – represents the measurable/observable utility for alternative j

β_{0j} – represents the average influence of all observed factors on utility

β_{kj} – represents the effect of a unit change in attribute k to utility

X_{kj} – represents the level of attribute k for choice j

Under this definition of utility β_{kj}, the effect of attribute k on choosing option j, is assumed to be the same for all respondents. In other words, for example, it assumes people's tastes for a given route characteristic do not vary across transport users.[23] This is an important limitation of logit models and is discussed further below. Each alternative will have a unique utility function, however for model identification purposes, the constant term, β_{0j}, for one alternative must be normalised to zero.

If we could measure directly each respondent's utility for a given option (V_j), then we could carry out a standard regression to find the coefficients (β_{kj}) that best fit the observed choices. However, only the respondents' choices are recorded in the data, not V_j itself. Thus, CM uses the actual choices made, and the attributes associated with the alternatives, to predict the probability of an individual choosing a particular alternative (Train 2002). The probability of an individual selecting alternative j over all other alternatives (J) in a choice set can be denoted as:

$$Prob_i = Prob\left[(V_j + e_j) \geq (V_J + e_J)\right] \qquad \text{(Equation 3.3)}$$

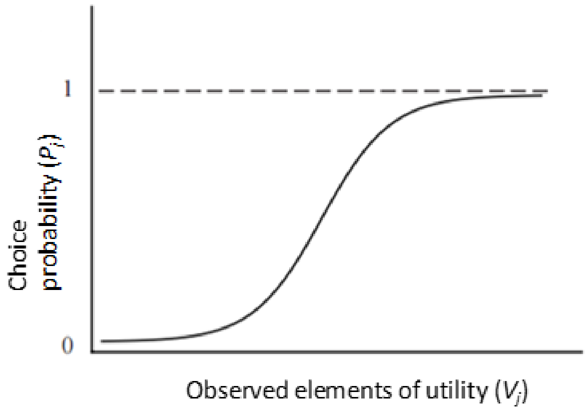
When predicting probabilities for separate (discrete) outcomes, such as those in a choice set, an appropriate regression model must be used. The logit and mixed logit models are commonly employed for this purpose (Scarpa and Rose 2008; Meade and Cheung 2016). These two techniques are discussed below.

[23] Holding constant the effect of demographic variables included in the explanation of utility.

3.4.1 Logit model

The logit (or multinomial logit or MNL) model converts the choice frequencies for a given alternative into choice probabilities via the logistic function. The logit probabilistic function is an 'S' shaped curve with the minimum and maximum limits of 0 and 1 used to model how the probability of an event may be affected by one or more explanatory variables (Train 2002). In figure 3.4 we see how the probability of an alternative being chosen increases with the level of utility obtained from the choice.

Figure 3.4 Logit model



Using the MNL model, the probability of a respondent selecting option j is the exponential of its observed elements of utility (V_j) divided by the sum of the exponential of observed utility for all options in the choice set (V_J).

$$Prob_j = \frac{exp^{v_j}}{\sum_J exp^{v_J}}$$ (Equation 3.4)

Best practice regression modelling requires all important explanatory variables to be included in the model. In this case, utility (V) from an alternative is a function of its attributes (ie Lancasterian theory) and respondent characteristics. Characteristics included as additional explanatory variables in CM often relate to a person's socioeconomic (eg wage/salary), attitudinal and demographic (eg age) status (Lee, 2012). The expanded version of the above logit probability equation is as follows, where β represents the effect of an attribute or individual characteristic (X) to choice probability, and β_0 represents average influence of all observed factors on choice probability.

$$Prob_j = \frac{exp^{\beta_{0j}+\beta_{kj}.X_{kj}}}{\sum_J exp^{\beta_{0J}+\beta_{kJ}.X_{kJ}}}$$ (Equation 3.5)

Estimation of each explanatory variable's effect on probability is achieved via the process of maximum likelihood estimation (MLE), which uses a repetitive algorithm to fit the MNL model that best explains the choices made by respondents.

3.4.2 Mixed logit model

As mentioned above, the standard logit model is limited by its assumption that taste or preference for a given attribute (represented by an attribute's regression coefficient, β) is the same for all respondents. However, transport studies commonly find tastes for different transport route/mode characteristics vary across travellers (Meade and Cheung 2016).[24] Therefore, transport choices may be better explained by the mixed logit model, a model that allows tastes to vary.

The mixed logit model redefines the measured element of utility, now V_{ji}, to account for individuals' (i) heterogeneous preferences for option j:

$$V_{ji} = \beta_{0ji} + \beta_{1ji}.X_{1j} + \beta_{2ji}.X_{2j} + \ldots + \beta_{kji}.X_{kj} \quad \text{(Equation 3.6)}$$

Because taste parameters, β_{kji}, now vary by respondent, the mixed logit model produces a probability equation that cannot be solved directly. Statistical techniques can be used to estimate taste parameters. However, a *distribution* of taste parameters is produced, rather than a single point estimate for a given attribute. Consequently, mixed logit WTP values are also distributions instead of point estimates. In short, the flexibility of the mixed logit model enables evaluators to understand how different types of travellers differ in their trade-offs between transport route attributes (Meade and Cheung 2016).

3.5 Specifying utility functions

The most commonly assumed utility specification is a linear utility function. While this is the most likely function to be estimated here, non-linear utility functions should be tested also. For example, it is theoretically possible the marginal utility (preference represented by the modelled parameter) for time, may differ between short and long trips. To test for this possibility, it may be necessary to impose non-linear transformations to the attributes at the time of estimate, eg taking the square of an attribute, its log or square root, or applying some form of non-linear categorical transformation, such as the application of dummy or effects coding.

It is important to note the final decision as to the use of non-linear transformations should be made either for:

- theoretical reasons, eg the analyst may have some specific economic rationale for assuming a specific non-linear transformation should be used, or
- empirical reasons, eg the analyst may discover a particular non-linear transformation provides a better model fit to the data. As discussed in the next section however, the use of non-linear transformations to the data may have ramifications on important model outputs.

A further point of clarification is necessary when dealing with SC experiments which use travel time distributions to capture the concept of travel time reliability. When specifying the utility function for this type of problem, the individual travel times and frequency of occurrences do not enter the utility function directly. Rather, the expected travel time and variance of travel time are computed, and it is these values that enter into the utility specification. It is necessary for the attributes shown to respondents in each SC task to be transformed into the variables of interest for the study, eg travel time (expected travel time) and reliability (the variance of the travel time distribution).

3.6 Model outcomes

Of primary interest for the current study are the marginal rates of substitution between the attributes used in the SC experiment. When cost is one of the attributes used, the marginal rate of substitution is known

[24] From a random utility theory perspective, this is a consequence of not being able to observe all factors that influence traveller's behaviour.

as WTP. The literature has identified two primary types of WTP measures that can be derived from models of discrete choice: marginal WTP and total WTP. Here, we are concerned only with marginal WTP.[25]

In transport analysis, the WTP for time, is commonly referred to as the VoT, the VTTS, or the subjective value of time (SVT). VoT, VTTS and SVT represent the same measure, ie the WTP for time, and hence may be used interchangeably. Preference for which terminology is used is typically based on the geographical location of the researcher (VoT is typically used in Europe, VTTS in Australasia and SVT in South America); VoT is used in the EEM. Transport decision makers are also increasingly interested in calculating the WTP to improve travel time reliability, which is often referred to as the value of travel time reliability (VTTR) or sometimes as VoR. The WTP for attributes other than time and reliability tend not to have pre-specified nomenclature within the transport literature.

Marginal WTP describes how much the cost attribute, x_c, would be required to change given a one unit change in an attribute, x_k, so the change in total utility will be zero. In other words, how much someone is willing to pay extra for an increase in another attribute. The marginal WTP is calculated by taking the ratio of the derivatives towards both the attribute of interest and cost, which in the case of a linear in the attributes indirect utility specification, is given by the following equation.

$$WTP_k = -\frac{\Delta x_c}{\Delta x_k} = -\frac{\frac{dV}{dx_k}}{\frac{dV}{dx_c}} = -\frac{\beta_k}{\beta_c}. \qquad \text{(Equation 3.7)}$$

To explain the negative sign in the equation above, note the cost parameter is typically negative. When attribute k has a positive parameter (eg effectiveness of treatment), then one is willing to pay more for a more effective treatment. In contrast, when $\beta_k < 0$ (eg when attribute k describes the amount of adverse side effects), then one is willing to pay more for fewer side effects.

The WTP to avoid injury or death is not the same as the VSL or value of statistical injury (VSI). The VSL (or VSI) is calculated as the WTP to avoid death (or injury) after adjusting for the risk of exposure of such an outcome. As such, while the WTP to avoid injury or death represent inputs into the VSL or VSI calculations, the WTP values are not the VSL or VSI values themselves.

[25] For a discussion of total WTP measures, see eg Lanscar and Savage 2004

4 Survey design

4.1 Introduction

The development of a CM survey is a multi-step process. In this section we discuss initial elements of a survey design for testing with small groups of respondents to obtain feedback.

SC experiments require the analyst to construct a set of hypothetical scenarios involving two or more options, at least one of which is described by a set of attributes and attribute levels. Respondents completing the survey are then asked to review these scenarios and indicate their preferred alternative based on the attribute levels shown. The allocation of the attribute levels to the survey task typically occurs via an experimental design, although random allocation is not uncommon in practice.

For any SC study, many potential experimental designs can be constructed and used. The final design(s) chosen will depend on a series of assumptions the analyst either makes deliberately, or does so without knowing. Before discussing these assumptions, we first discuss decisions that must be made prior to the design being generated. These decisions are independent of the design generation process used, but they may influence the design process significantly.

4.2 Design dimensions

The SC tasks shown to respondents are characterised along several dimensions, including the number of alternatives, number of attributes, attribute levels, the range of the attribute levels, as well as the number of tasks respondents are asked to complete. Here we provide a brief description of how each of these dimensions can affect the results of the experiment.

4.2.1 Response mechanism

4.2.1.1 Context clarity

It is critical the description and explanation of question contexts and all available alternatives are absolutely clear. Respondents must be fully informed of what they are valuing or comparing, in terms of both quantity and quality (attributes), for results to be valid (Moller 2012).

Respondents may not have considered, or may not have sufficient pre-existing knowledge about given alternatives to select their most preferred option without first being provided all relevant information they require. This is particularly important for research investigating preferences around prospective upgrades, as given alternatives may be associated with potential implications respondents are unlikely to consider without being overtly informed first (Moller 2012).

4.2.1.2 Response type

SP studies can be designed to allow for the respondent to answer questions in several ways. These vary from simple (for example, selecting one preferred option from two alternatives), to complex (for example, ranking five alternatives from most to least preferred). Each response type allows for a certain type of data and results to be drawn from the study.

Louviere et al (2000) suggest, to mimic reality, surveys should ask respondents to select their single most preferred option from two or more alternatives. They reason it is not appropriate to ask respondents to rank alternatives from most to least preferred, as this is generally not how people go about making decisions on a day-to-day basis, and this may lead them to respond in a way they would not in reality. Similarly, Hensher et al (2005) stress the importance of designing a survey that aligns with reality as much as possible to produce robust and accurate results. Another benefit of using such a simple response form is the likely reduction it would have on survey complexity and length (Louviere et al 2000)

However, Hess and Rose (2009) acknowledge using more in-depth response forms, such as asking respondents to rank alternatives from most to least preferred, selecting their best and worst alternatives or estimating how frequently they may use each alternative when provided with a fixed set of 'uses', may provide richer and more informative data. These results therefore have the potential to go beyond a singular preference to reveal further information into how respondents value alternatives relative to one another.

Accent and RAND Europe (2010) suggest a technique which achieves the benefits of both single selection and preference ranking whereby respondents are initially asked to select their single most preferred alternative; then in a sequential question they are asked to rank the remaining alternatives from their most to least preferred. Accent and RAND Europe (2010) also recommend either a 'status quo' or 'none of the alternatives given' option be provided, so respondents are not forced to select an option they may not want for the sake of answering the question.

In this study, we concentrate on a response mechanism that captures first preference rankings only via a 'pick one' response, although the theories discussed can easily be extended to other response mechanisms. For example, recent developments in SC research have seen an increase in obtaining either partial or full rankings of data via, say, a best/worst response (eg Louviere et al 2015). Aside from psychological issues surrounding the response mechanisms, the type of response and the number of options shown may affect the dimensions of the design, including the minimum number of choice tasks required for estimation purposes.

4.2.2 Labelled or unlabelled

One important consideration is whether the experiment will be *labelled* or *unlabelled*.

- *Labelled experiments* involve tasks where the names of the alternatives shown to respondents have some substantive meaning beyond the relative order of appearance in the task, eg car, bus, train.
- *Unlabelled experiments* are where the names of the alternatives convey only their relative order of appearance in the task, eg Policy A, Policy B, Policy C.

From the perspective of design generation, this decision is important as it might directly affect the type and number of parameters that can or will be estimated as part of the study. Typically, unlabelled experiments will involve the estimation of generic parameters whereas labelled experiments may involve the estimation of option-specific and/or generic parameter estimates, hence potentially resulting in more parameter estimates than with an identical, though unlabelled, experiment. An example is where travel time has a different value depending on the type of environment the route traverses. The labels 'country route' and 'urban route' would allow differentiation of this effect. Without the labels, time is a generic attribute with the same value in any choice alternative. The labels may also convey other unstated information to respondents which influences their choices. A difficulty for the analyst is they are unaware of what additional label-signalled information is processed by each respondent.

This issue became relevant in this study during the alpha and beta testing (see chapter 5) when maps were used to illustrate the alternatives in a choice set. Study participants made us of the ('labelled') information in the maps but this was not controlled in a way which allowed us to measure the influence on the choices made. In the pilot study the maps were dropped.

4.2.3 Dimensions

SP studies ask respondents to select their most preferred option among two or more alternatives (for example, different transport routes) each with associated attributes (eg travel time, safety, scenic qualities) occurring at different levels (eg travel time taking 10 minutes, 15 minutes or 20 minutes). When designing a SP study, it is important to consider how much information is provided to respondents in terms of alternatives, attributes and attribute levels, so they have enough information to make an informed decision but not so much they are overwhelmed and/or fail to consider all important information. Table 4.1 describes the design dimensions central to SP surveys. Below we discuss the different elements in turn.

Table 4.1 Dimensions of a SP study

Dimension	Description
Choice tasks	The number of choice questions offered, eg one choice question would be do you prefer route A or route B
Alternatives	The number of options to select from in each choice task. In this study, each choice offers two alternatives (route A or B)
Attributes	The number of characteristics for each alternative, eg average travel time, lateness, congestion
Attribute levels	Variations in the measure of an attribute offered across the alternatives

4.2.4 Number of choice tasks

The *number of choice tasks* is limited in practical terms by the concentration threshold of participants and, statistically it is bounded from below, by the number of degrees of freedom,[26] which is influenced by the number of parameters to be estimated, the number of alternatives to be shown to respondents, as well as by other considerations such as the number of choice tasks required to achieve attribute level balance. Additionally, assumptions and/or restrictions made by the analyst in generating the design may restrict the number of choice tasks.[27]

Beck et al (2011) report no differences in response rates and no systematic differences in the respondents' self-reported perception of the uncertainty as the number of choice tasks they were asked to complete increased. They found minor differences in mean WTP, but no differences in WTP standard deviations. Brazell and Louviere (1998) tested the effect of varying the number of choice tasks between 16 and 120. They found the variance in models estimated from statistically equivalent sub-designs decreases initially with increases in number of choices (learning), then goes up (fatigue). Caussade et al (2005) found the number of choice sets appears to have a relatively small influence on mean VTTS, a finding supported by Hensher (2004). Hensher (2006), however, did find the number of choice tasks impacted on the error variances, with a small but significant increase in error variance when the number of choices increased from six to fifteen in a six attribute choice experiment. Similarly, Rose et al (2009) found the number of

[26] The number of pieces of data that are free to vary.

[27] A (fractional factorial) orthogonal design (see below) sometimes needs (many) more choice tasks than the minimum number determined by the number of degrees of freedom and attribute level balance, merely because an orthogonal design may not exist or may be unknown for these dimensions. A full factorial design has a predetermined number of choice tasks, only influenced by the total number of attributes and the number of attribute levels.

choice tasks had almost no impact for an Australian data set, a limited impact on a Taiwanese sample, and a very large impact on a Chilean sample.

4.2.5 Number of alternatives

Vanniyasingam et al (2016) completed a study which investigated the optimal number of alternatives and corresponding attributes which can be asked about in each question. They acknowledge it is not feasible to present many options at one time, each with many varying attributes, because of the cognitive burden it places on respondents. Instead, select attributes and alternatives should be asked about successively over consecutive questions to gradually establish which attributes and alternatives respondents most value relative to one another.

They found survey efficiency increased as the number of alternatives provided increased and the number of attributes decreased. This suggests designs with many alternatives, each with a small number of attributes, performed the best. Designs with binary attributes produced results with the highest statistical significance compared with those with three or more associated attributes with each alternative.

Hensher (2001a) also investigated how much information in the form of alternatives and attributes respondents can handle in a SP survey design. Consistent with Simon (1955), he suggests, when respondents become too overloaded with information they begin to ignore some attributes altogether and make decisions based on select information only. Hensher (2001a) concluded that the extent to which respondents ignore attributes increases as the number of levels of each attribute increases, as the range of each attribute narrows and as the number of alternatives decreases. These results endorse the use of SP surveys with many alternatives, each associated with few attributes divided into few levels.

Adamowicz et al (2005) found a three-alternative version resulted in a higher probability of the status quo alternative being chosen than a two-alternative version of the same survey. Rolfe and Bennett (2009) also acknowledge the number of alternatives provided in each choice set is critical when it comes to survey design. They completed a study to test the optimal number of alternatives SP studies should offer. This was done by splitting a sample randomly into two groups, with one group receiving two alternatives per choice set, while the other received three alternatives. They found more robust results were produced from the group provided with three options. It appeared, when asked to choose between two options, respondents were more likely to engage in ‘serial non-participation’ or repeatedly selecting the same alternative without taking its adjusted attributes into account. Perhaps offering more than two options prompted respondents to engage in a more complete reasoning process. As a result, Rolfe and Bennett (2009) suggest SP studies should offer more than two options for respondents, and should also include a status quo option for comparison.

However, DeShazo and Fermo (2002) found, as the number of options offered increases, so does survey complexity and choice inconsistency. Considering these results together, it seems likely that, when designing a SP survey, enough options should be offered to allow for full participation and reasoning between options, but not so many the cognitive burden interferes with reasoning ability.

Several studies have examined error variance, ie the model’s ability to estimate respondents’ choices. Arentze et al (2003) looked at how the error variances between SC experiments with two and three alternatives might differ and found no differences to exist. DeShazo and Fermo (2002) found a quadratic relationship between the number of alternatives and the variance, suggesting error variance first decreases, then increases with the number of alternatives. Caussade et al (2005) also examined the impact of increasing the number of alternatives upon error variance. They concluded the number of alternatives used had the second largest influence on error variances out of all possible design dimensions – with four alternatives being better than three or five in terms of scale effects.

The literature is not conclusive on this issue, and the realism of having more than two routes available is highly unlikely. We have used an experimental design involving two alternatives in each choice set.

4.2.6 Number of attributes

Green and Srinivasan (1990) argue people cannot process many attributes at once; they become tired and consequently ignore or address attributes in random and uncontrolled ways, or tend to use heuristics that lead to biased preference measures. Both Arentze et al (2003) and Caussade et al (2005) found significant impacts on error variance as the number of attributes in the SC task increases. Arentze et al (2003) found increasing from three to five attributes led to increased error variances and parameter differences. Caussade et al (2005) concluded the number of attributes used had the largest influence on error variances out of all design dimensions. A similar finding was reported by DeShazo and Fermo (2002) who found, on average, an increase in the number of attributes increases the variance of the error component in utility.

Hensher (2006) and Rose et al (2009) found the number of attributes has a significant influence on parameter outputs and WTP measures, while Rose et al (2009) found possible cultural differences, with a downward bias being observed for Australian and Taiwanese respondents and an upward bias for a Chilean dataset as the number of attributes increases. The conclusion is, the number of attributes should be minimised to provide least variance in parameter estimates, and there are limits to people's cognitive ability to handle more than a few attributes concurrently. However, where it is desired to value several attributes, it is efficient to do so within a single experimental setting.

4.2.7 Number of attribute levels

Caussade et al (2005) describe how the number of attribute levels had the second least impact on error variance out of all design dimensions tested. Research by van der Waerden et al (2004), however, suggests the number of attribute levels may have a significant effect on estimated utilities and hence choice probabilities. Similarly, Hensher (2006) found the number of attribute levels affects the probability of respondents not ignoring an attribute in some, but not all cases. Rose et al (2009) found the number of attribute levels appears to have a significant impact on results, with differences depending on which country the data was from.

The *number of attribute levels* used is an important design decision, as is how they are to be modelled. This will influence the number of parameters that can be estimated. If nonlinear effects are expected for a certain attribute, then more than two levels need to be used for this attribute to be able to estimate these nonlinearities. Typically, dummy and/or effects coding is used to estimate these nonlinear effects. In this case, the number of parameters needed for each attribute is the number of levels minus one. Hence, the more levels used, the higher the minimum number of choice tasks required will be, as each additional level will require an additional parameter to be estimated. However, if the attribute is a continuous variable, one can also decide to estimate linear effects with only a single parameter per attribute, even though in the design multiple levels are used. In this case, the minimum number of required choice tasks does not increase as the number of levels increases.

Another decision required prior to generating the design is whether *attribute level balance* will be imposed when generating the design. Attribute level balance occurs when each attribute level appears an equal number of times for each attribute over the design. The imposition of attribute level balance may in turn affect the size of the final design. Further, mixing the number of attribute levels for different attributes may yield a higher number of choice tasks or alternatives required for a given design (due to attribute level balance). For example, if there are three attributes with two, three and five levels, respectively, then the minimum number of choice tasks for an attribute level balanced design will be 30 (since this is divisible by two, three and five without remainder). However, if one would use two, four and six levels, then a minimum of only 12 choice tasks would be required. Therefore, it is often suggested not to mix designs with too many different numbers of attribute levels, or at least have all even or all odd numbers of attribute levels, if attribute level balance is adjudged to be a desirable design criterion to impose during the design generation process.

4.2.7.1 Orthogonal designs

A simple design approach is full factorial design, where choice tasks cover all possible combinations of attributes and levels. Such designs have desirable qualities, including:

- balance – each level in each attribute appears the same number of times in the design; and
- orthogonality – all possible pairs of attributes appear the same number of times in the design.

Generally, an orthogonal and balanced design can help produce uncorrelated parameter estimates (eg high values of one parameter, such as trip length, are not frequently found with high values of another, such as crash risk) and precise (eg estimates with small standard errors), respectively (Mangham et al 2009). However, full factorial designs can be impractical, as the number of choice sets grows rapidly in size with the number of attributes and levels. Fractional orthogonal designs or more simply 'orthogonal' designs – a method which selects a subset (fraction) of all possible choice sets, yet maintains the orthogonality and balancing properties of a full design – were the traditional solution to this problem (De Dios Ortuzar and Willumsen 2011).

Table 4.2 provides an example of a basic survey design consisting of three attributes (price, safety and time) with two levels each (one or two). In this case, a full factorial design would require eight scenarios, and the minimum fractional orthogonal design would require four scenarios – the minimum number of scenarios with each possible attribute pair for price and safety, price and time and safety and time is presented (eg one and one, one and two, two and one, and two and two). The non-orthogonal subset gives rise to correlated attributes, for instance, level two for time is found more frequently with level two for safety than level one for safety. This is problematic as linear regression models cannot accurately estimate the effects of highly correlated variables (attributes).

Table 4.2 Example of a full factorial, orthogonal factorial and non-orthogonal factorial designs

Full factorial				Orthogonal fractional				Non-orthogonal fractional			
Scenario	**Price**	**Safety**	**Time**	**Scenario**	**Price**	**Safety**	**Time**	**Scenario**	**Price**	**Safety**	**Time**
1	1	1	2	**1**	1	1	2	**2**	1	1	1
2	1	1	1	**4**	1	2	1	**4**	1	2	1
3	1	2	2	**6**	2	1	1	**3**	1	2	2
4	1	2	1	**7**	2	2	2	**7**	2	2	2
5	2	1	2								
6	2	1	1								
7	2	2	2								
8	2	2	1								

Attribute correlations								
Price-safety	**Safety-time**	**Price-time**	**Price-safety**	**Safety-time**	**Price-time**	**Price safety**	**Safety-time**	**Price-time**
0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.58	0.58

Source: Sanko (2001), modified version of table 4-2-1

4.2.7.2 Blocking

In cases where the optimal number of orthogonal choice tasks generated for a SP study exceeds the number of choice tasks deemed reasonable, a technique known as 'blocking' may be employed. Bliemer et al (2009) describe blocking as allocating a certain subset of choice tasks within the survey to each respondent. This technique allows data to be gathered about all choice tasks of interest in an efficient manner, while reducing the cognitive burden placed on respondents.

Bliemer et al (2009) suggest blocking should not be randomised, but instead allocated formally to ensure each respondent is provided an even and representative set of choice tasks. As orthogonal designs tend to have a relatively small set of choice tasks, randomly assigning these to respondents may mean each respondent will be provided a set of choice tasks differing significantly from that provided to the other respondents. As a result, respondents may provide responses unequally influenced by the random set of choice tasks they had been allocated. Hess and Rose (2009) found a systematically blocked orthogonal design performed significantly better than a randomly blocked orthogonal design, and only slightly worse than a much more sophisticated efficient design.

4.2.8 Design efficiency

Seeking attribute level balance can lead to larger than necessary surveys. Furthermore, non-linear techniques, such as the logit model, are now the preferred evaluation tool for SP data and do not rely on orthogonality to derive robust estimates (Hess and Rose 2009; Scarpa and Rose 2008).

Consequently, emphasis has shifted from orthogonal designs to design efficiency: minimising the sample size and attribute/level combinations necessary to achieve a given degree of estimation accuracy (Scarpa and Rose 2008). Essentially, this notion reflects the analysts' desire to find reliable estimates at least cost. Large samples are expensive, particularly if interviews are required, and lengthy surveys induce questionnaire fatigue (see section 4.5.5) which leads to estimation inaccuracy (Johnson et al 2013). Indeed, empirical evidence suggests efficient designs produce more robust estimates than those from orthogonal designs (Rose and Bliemer 2012).

Several measures have been proposed to test design efficiency, including:

- A-efficiency – minimises the sum of variances of an individual parameter estimate (Kim and Haab 2004)
- D-efficiency – minimises the mean variance around multiple estimated parameters (Kim and Haab 2004)
- C- efficiency – minimises the variance around the summary statistic of interest, eg WTP estimate
- G-efficiency – compares the estimated standard errors against the minimum possible standard error from a full factorial design (Rao 2014)
- S-efficiency – minimises the sample size requirements for a given the set of attributes and levels

Scarpa and Rose (2008) note the study objectives determine which measures are more appropriate. For example, the D-efficiency criterion will primarily help minimise standard errors and estimate covariances, while the S-efficiency may be preferred to reduce sample size and survey cost. However, if the aim is to examine WTP values, Scarpa and Rose (2008) recommend the C-error criterion, as it is tailored to optimise designs centred on WTP estimates.

Complicating the search for an efficient design is the information required to calculate the standard errors around the final parameter estimates. This requires prior knowledge of the parameter values of interest, otherwise known as 'priors'. These values are unknown; this uncertainty is the reason for the survey. At this point, two different approaches are taken by the literature. The first approach seeks to minimise the elements of the standard errors by assuming the parameters will be zero (ie uses priors equal to zero) and the attributes will be orthogonally coded (Bliemer et al 2008). The second approach assumes researchers have some idea about the direction and size of parameter values (through the literature, pilot studies, focus groups or expert judgement), and thus minimise standard errors using non-zero parameter values (Scarpa and Rose 2008). Originally, researchers employing non-zero priors (for example, assuming drivers are willing to trade three minutes of travel time for $1) used exact estimates of a given parameter; however, recent studies employ *Bayesian efficient designs*, which allow the true parameter to fall within some distribution of values (De Dios Ortuzar and Willumsen 2011). The main benefit of non-zero priors is the design is directly related to the expected outcome, which helps reduce the variance of estimates, ie improve design efficiency. We use non-zero priors in this study, building on existing data and the alpha and beta test experience.

Once an efficiency measure and prior has been decided on, the final task of design generation is to search for arrangements of attributes and attribute levels best satisfying the chosen criterion. Computational-intensive algorithms are required for this step. Common techniques include the:

- Modified Federov algorithm – generates a large number of choice sets from which a smaller number are randomly selected. The choice set combination with the lowest D-error is kept as the most efficient design.
- RSC algorithm – generates a single design, and then rearranges attribute level combinations via three different processes: *relabelling*, *swapping* and *cycling*. The algorithm alternates between these processes until no efficiency improvements are possible.
- Random exchange algorithm – generates several random designs from which the best design is chosen according to a given efficiency criterion. Then, within a single attribute, levels are shuffled until no efficiency improvement can be made. This is process repeated for each attribute.

Quan et al (2011) found ‘random exchange’ generates more efficient designs, and in less time, than other methods, including the modified Federov and RSC algorithms.

4.2.9 Attribute level range

The *attribute level range* of quantitative attributes is another important decision needing to be made before the experimental design can be determined.

Several researchers have explored how changing the range used to describe attributes can influence the outcomes of discrete CMs. Ohler et al (2000) found attribute range differences affect experimental outcomes in terms of complexity of functional forms, model fit, power to detect non-additivity, and between-subject response variability; no effect was found on model parameters, within-subject response variability and error variance. Contrary to this finding, Caussade et al (2005) concluded that the range of attribute levels had the third largest influence on error variances out of all design dimensions. With respect to the value of time, Hensher (2004) found wider ranges resulted in a lowering of the mean VTTS compared with using a narrower range for attributes.

Research suggests using a wide range (eg $1–$6) is statistically preferable to using a narrow range (eg $3–$4) as this will theoretically lead to better parameter estimates (ie parameter estimates with a smaller standard error), although using too wide a range may also be problematic (see Bliemer and Rose 2010). This is because the attribute level range will affect the likely choice probabilities obtained from the design, which will affect the expected standard errors from the design. Having too wide a range may result in choice tasks with dominated alternatives (at least for some attributes), whereas too narrow a range will result in alternatives that are largely indistinguishable from one another. This is a pure statistical property and one should take into account the practical limitations of the attribute levels, in so much as the attribute levels shown to the respondents have to make sense. Therefore, there is a trade-off between the statistical preference for a wide range and practical considerations limiting the range.

4.2.10 Adaptive conjoint analysis

It may be useful to incorporate adaptive conjoint analysis (ACA) into a SP study which requires the inclusion of many attributes (and attribute levels). ACA used by Sawtooth Software is a popular package used to predict preference judgements and choice behaviour (Argarwal and Green 1991). Its creators define ACA as a 'computer-administered, interactive conjoint method designed for situations in which the number of attributes exceeds what can reasonably be done with more traditional methods' (Sawtooth 2015, p1). ACA customises each respondent's experience to narrow in on the attributes (and attribute levels) they consider the most important when making trade-offs. As a result, ACA avoids overloading respondents with too many attributes at any one time (Sawtooth 2015). However, ACA suffers from possible statistical biases such as endogeneity biases, as subsequent questions in the survey are based on previously observed results, meaning the error terms across choice tasks may be highly correlated.

ACA relies on transitivity of preferences, which assumes the relation between attributes holds when different alternative combinations are provided. However, McNamara et al (2014) suggest, in certain situations, individuals may make seemingly irrational decisions which conflict with the assumption of transitivity. This is a naturally occurring phenomenon which does not necessarily indicate individuals have made an error of judgement. McNamara et al (2014) suggest it may be natural to make decisions that violate transitivity when it is possible a certain alternative may disappear in the future, or conversely a better alternative may appear. As such, SP studies relying on transitivity of preferences should be mindful of the impact removing/adding alternatives to successive choice tasks may have.

4.2.11 Qualitative research

Although SP studies are traditionally quantitative, there is value in using qualitative research elements, particularly at the pilot testing stage, eg asking respondents why they chose one alternative over another and how they reached their decision. Moller (2012) cites the potential usefulness of this exercise to more deeply understand respondents' reasoning processes.

Klojgaard et al (2012) emphasise incorporating qualitative research to inform the design of a quantitative SP research project. They acknowledge the complexities associated with CM survey designs and recommend using qualitative research to optimise the design of a successive quantitative survey. They suggest the use of a literature review, observational fieldwork, interviews with potential respondents and a qualitative pilot test (including cognitive testing) before undertaking a large scale quantitative SP study. In particular, they suggest using qualitative methods to test how respondents understand and consider situational context along with given alternatives and their respective attributes. This may reveal factors respondents consider when making decisions, which had not previously been considered or included in the SP survey design. It also may identify what attributes are most, least, or not at all important to respondents when they make such decisions.

4.2.12 Conclusions

There is no clear indication of exactly how many alternatives, attributes and attribute levels are optimal to include in a SP survey, as this depends on the complexity of the survey context itself. The studies discussed above indicate respondents tend to be able to handle many alternatives, but may suffer from too heavy a cognitive load when many attributes and attribute levels are associated with each alternative.

Pilot testing can be used to better understand how well respondents handle questions with varying amounts of information, particularly the number of attributes described (DeShazo and Fermo 2002). If respondents are providing inconsistent preferences across equivalent questions, it may indicate they are failing to take all information into account and the number of attributes and/or attribute levels should be decreased. To maintain balance throughout the survey, attribute levels should occur with an equal frequency within a pre-determined optimum range.

4.3 Attributes of interest

4.3.1 Value of safety

Many researchers have expressed difficulty when asking respondents to make trade-offs between safety risk and monetary value. This is attributed to the trouble respondents have internalising and processing relatively small risk probabilities. Rizzi and Ortuzar (2006) suggest it is not useful to include safety by providing the probability a journey will produce a fatal crash. These probabilities (even when relatively large) tend to be too small for respondents to conceptualise and compare. It has also been suggested respondents often struggle to understand and interpret probability figures.

Rizzi and Ortuzar (2006) tested the robustness of a SP study to evaluate how well respondents could make sense of travel safety. They found more consistent and reliable results when respondents were told how many fatal crashes are expected to occur over a certain time period, rather than the probability of a fatal crash occurring on any single journey. Rizzi and Ortuzar (2006) found many respondents tended to rank alternatives from most to least safe based on how many fatalities occur in a predetermined time frame to simplify more complex information they had been provided. Respondents then used this information in making their choice.

Pilot testing is used in this study to gauge potential respondents' understanding of safety risk, and to ensure respondents can fully understand and conceptualise safety risk in the form it is presented. One difficulty with the Rizzi and Ortuzar (2006) 'risk per unit of time' framing of safety risks is it does not convey the risk, absolute or otherwise, to the individual traveller without information on exposure, ie the volume of traffic affects risk. In the present study this was addressed by expressing the risk in terms of billions of vehicle kilometres travelled, resulting in single digit measures applicable to an individual trip. Differences in risk are highlighted by comparison with mean national risk using the same units.

4.3.2 Time

CM, particularly SP studies, appear to be a particularly favoured method of valuing travel time savings in much of the literature. This is because individuals tend to find assigning monetary values to travel time savings difficult. Often, assigning value to travel time savings is not something individuals have considered before, so when asked to do this they are more likely to indicate a value with a low level of reasoning behind it, or simply refuse to answer (Accent and RAND Europe 2010).

As discussed in section 2.3.1, studies have found respondents can experience difficulties in valuing small time savings. Respondents also tend to value travel time savings in the context of their overall journey duration. Wallis et al (2015) found international evidence that unit values (per minute) for longer distance trips (3+ hours) are 50 –100% greater than values for shorter trips (<20 minutes). However, as noted in chapter 2, the literature is not consistent on this point. Metz (2008) suggests saving five minutes on a trip expected to take twenty minutes is likely to be considered a valuable saving, but if the trip is expected to take two hours, it is likely to be considered less valuable. Similarly, Wallis et al (2015) note 'the UK literature in particular refers to the “proportionality” argument, which would result in unit values decreasing with distance or duration’. This discrepancy is beyond the scope of this study.

4.3.2.1 Travel time for tourists

Tourists have different trip motivations and available time patterns and therefore a unique VTTS, compared with other transport network users. Nicolau and Más (2006) suggested the distance, cost and travel time associated with visiting a location is moderated by the strength of motivation the tourist has to visit the location. Tourists are willing to travel further, for longer to visit locations which they are highly motivated to visit, compared with non-tourists; they can have more time available as they are on holiday.

Schroeder and Louviere (1999) investigated how tourists visiting recreational sites responded to fees of various types and levels within a SP study. They suggest tourists tend to be willing to invest greater travel time to reach a destination they wish to visit for a longer period of time. These studies demonstrate the unique nature of VTTS associated with tourism–based travel. If tourists are to be included in a transport SP study, it would be useful to include destination attributes such as level of motivation to visit a certain destination, and the duration of stay at the destination, alongside other standard transport attributes.

Since tourists have differing priorities and motivations they require separate sampling and analysis form the general population.

4.3.3 Reliability

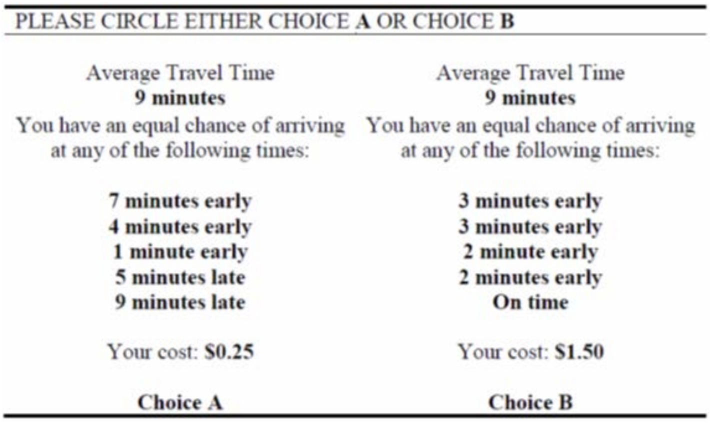
Travel times tend to vary for the same journey. To include time as a realistic attribute in a SP study, Hess and Rose (2009) suggest including time variability as a measure of reliability. Hess and Rose (2009) recommend time variability is included alongside total travel time as a semi-attached attribute in the form of +/- minutes. For example, providing journey duration in the form of X minutes +/- Y minutes (to produce a more realistic journey duration range). Hensher (2001) supports this suggestion, adding that incorporating time variability in the survey increases the richness and usefulness of results.

Choice modelling approaches to measuring the value of reliability (VOR) have varied and become more sophisticated over time. Initially simple choices were given, consistent with the mean variance model (see section 2.4); travel time variability was presented as the extent and frequency of delay relative to normal travel time. Li et al (2010) call these approaches ‘type 1’; they provide an example in which respondents are asked to make a choice between a journey which always takes 30 minutes and a journey which has a shorter time, but a possibility of five-minute delay once a week. But they note it is difficult for respondents to fully understand and interpret the travel time distribution from it.

More recently, multiple arrival times are given in choice experiments. These ‘type 2’ models provide a more sophisticated experiment. An example is given in figure 4.1 which incorporates both the mean variance model and the scheduling model approach; the design attributes in this experiment are mean travel time, travel cost, departure time shift, and standard deviation of travel time, while the attributes shown to respondents are mean travel time, travel cost, and five equally probable arrival scenarios (early, late or on time) with respect to the preferred arrival time.

These examples appear to be quite complex for the respondent, although we would expect the example to be interpreted as showing two options with the same travel time, one of which is lower cost but has greater travel time variability.

Figure 4.1 Type 2 choice experiment example


PLEASE CIRCLE EITHER CHOICE **A** OR CHOICE **B**

Choice A	Choice B
Average Travel Time **9 minutes**	Average Travel Time **9 minutes**
You have an equal chance of arriving at any of the following times:	You have an equal chance of arriving at any of the following times:
7 minutes early	**3 minutes early**
4 minutes early	**3 minutes early**
1 minute early	**2 minute early**
5 minutes late	**2 minutes early**
9 minutes late	**On time**
Your cost: **$0.25**	Your cost: **$1.50**
Choice A	**Choice B**

Source: Small et al (1999) in Li et al (2010)

The value of time reliability is not symmetrical around the mean, people value being early (compared with expected time) differently from being late (Small 1982; Oxera and Mott MacDonald 2003).

A second example (figure 4.2) asks respondents to choose between two options which have five equally possible trip characteristics, specified in terms of lateness of departure and arrival. It is a more complicated choice, without clear differentiation between the two options. Noting these complexities, for greater simplicity in the survey design, we have included only minutes late as an attribute and with less detail on the distribution of lateness (see chapters 5 and 6).

Figure 4.2 Example of mean lateness model

Source: Batley and Ibáñez (2009) in Li et al (2010)

4.4 Establishing functional relationships

The utility function, an equation of variables used to explain people's transport preferences, is central to CM of travel behaviour. Explanatory variables in the utility function correspond to attributes within the choice set or respondent characteristics (see section 3.3). Although the sensitivity of SP estimates to different utility function specifications has been long understood (De Dios Ortuzar et al 1997; Algers et al 1998), it has received relatively little attention from the choice modelling literature (Van der Pol et al 2014). Inconsistency of CM estimates is often attributed to systematic variation in transport preferences (tastes) not recognised by the CM (Hess et al 2005). Heterogeneous tastes can be accounted for by introducing interaction terms of traveller attributes (eg age or income) with route attributes (eg travel time or road safety), and applying a mixed logit CM which estimates individual–specific tastes for transport attributes (rather than an average 'taste' for the whole sample, see section 3.4.2). The former technique controls for observed heterogeneity while the latter accounts for unobservable heterogeneity (Carrion and Levinson 2012). When applying the mixed logit model, Meade and Cheung (2016) suggest not all taste parameters should be set as random. For instance, as travel cost is the common denominator of all WTP estimates, it should be set as constant across travellers so it can be accurately defined.

The literature is fairly inconclusive about the optimal range of attributes for modelling of traveller preferences. Hess et al (2005) suggest researchers include as many descriptive attributes as possible to reduce the impact of unobserved heterogeneity. However, such attempts to extract increased behavioural realism require extremely rich datasets, and may lead to mis-inference if data is of insufficient quality (Hensher and Greene 2003). Furthermore, respondents are unable to take all into account when deciding among options described by a large number of attributes (Lindhjem and Navrud 2011). Problems from burdening respondents with a complex survey are summarised in section 4.5.2.

4.5 Issues in survey design

The following summarises literature as observed phenomenon relevant to SP surveys.

4.5.1 Realism and hypothetical bias

When respondents are asked to select from one or more hypothetical alternatives, their responses may be different from how they would act in a similar real life situation. This 'hypothetical bias' can be quantified as the discrepancy between a respondent's revealed and stated preference (Loomis 2011). Many studies have documented the effect of hypothetical bias within SP studies (Moller 2012), along with how to quantify and minimise its effects (Accent and RAND Europe 2010). However, it is still unclear why there is such a discrepancy (Loomis 2011; Mitani and Flores 2014).

Hypothetical bias tends to have its most significant effect when contexts and alternatives deviate significantly from reality, as respondents are thought to have difficulty imagining and relating to the situation (Accent and RAND Europe 2010). Respondents tend to overestimate the value of goods when hypothetical bias is at play (Moller 2012), with estimates ranging from 30% more (Lusk and Schroeder 2004) to a factor of two to three times (Murphy et al 2005). Murphy et al (2005) suggest hypothetical bias may have a particularly large impact on SP studies which ask respondents to indicate their WTP for public goods.

Loomis (2011) has suggested use of careful survey design and/or post-field work recalibration to minimise the influence of hypothetical bias. First, he suggests creating a survey grounded as close to reality as possible: a life-like, easily imaginable, understandable and relatable scenario. This allows respondents to fully grasp the situational context and engage with a decision-making process similar to what they would use in a comparable real-life situation.

Loomis (2011) also suggests providing a 'cheap talk' script to respondents before they complete the survey. This script informs respondents that people tend to overestimate valuations (particularly for public goods) and simply asks them to be aware of this. This script encourages respondents to report their choice preference or WTP if the situation was one they were facing in real life. Many studies have found, when cheap talk scripts are used, respondents tend to provide lower WTP estimates (Carlsson and Martinsson 2006). While Mahieu et al (2012) attribute this to the minimisation of hypothetical bias as a result of the cheap talk script, it is possible respondents have provided deflated values for no other reason than because they had been encouraged to by the cheap talk script.

Finally, Loomis (2011) suggests recalibrating results to correct estimated value figures, such as WTP, after all data has been collected. This is done by deflating an individual's WTP by a constant factor. Tailored calibration based on a pilot study (compared stated and revealed WTP) is thought to be the most accurate method for calculating the extent to which hypothetical bias should be adjusted for.

4.5.2 Survey complexity and satisficing

SP surveys may challenge participants to engage in a deeper level of deliberation, reasoning and personal effort than is typically seen with traditional surveys (Mitchell and Carson 1989). Krosnic (1991) suggests, when asked to complete a complex survey, respondents may fail to put in the effort required to make adequately thought-out choices. Instead, respondents may short cut the reasoning process and offer inaccurate responses. Lindhjem and Navrud (2011) refer to this process as 'satisficing' and suggest the effect of satisficing tends to increase as the questionnaire becomes more taxing (in terms of both complexity and length); it also depends on how motivated and able the respondent is to participate in the research. As such it is important to refrain from assuming respondents will make decisions based on well thought out rational logic in which all alternatives are considered fully and are weighed up to make the best possible choice (Simon 1955).

Hess and Rose (2009) suggest satisficing occurs when respondents are provided with so much information they become unable to take it all into account when deciding. In this scenario, respondents may resort to reductive behaviour to overcome the complexity they face. This could include burden simplification and prioritising strategies such as throwing an attribute, or attributes, out to make choices easier, or more extreme behaviours such as mental 'coin flipping', selecting the first appropriate answer, acquiescence, endorsing the status quo, non-differentiating (offering same response repeatedly on different questions), selecting mid-point on ranking scales or selecting 'don't know' options (Carlsson 2009).

Carlsson et al (2012) found, as survey complexity increases, so too does choice inconsistency. This effect may be partially attributed to respondents resorting to satisficing strategies as survey complexity increases.

Using eye tracking technology Pieters and Warlop (1999) found, the more time respondents spent looking at certain alternatives, the more likely they were to select those alternatives as their choice. As such, it appeared the share of visual attention each alternative received predicted the share of times an alternative was selected as a choice. Alternatives which were skipped over or left unevaluated were therefore less likely to be selected due to the role of satisficing. This indicates the importance of understanding why, how and when respondents satisfice.

Simon (1955) suggested satisficing depends on the order in which alternatives are evaluated, with the first 'good enough' option being selected. Participants simply do not see the need to consider other alternatives after coming across a satisfactory alternative. This halt in searching is known as the 'stopping rule'. As a result, a satisficing respondent's choice is dependent on the order in which they evaluate each alternative. Stüttgen et al (2012) tested this model and found considerable support, particularly when personal consequences of the decision would have minimal personal effect (likely to be the case in hypothetic stated preferences survey responses). As a result, it is suggested the order of alternatives be randomised to minimise the effect of this form of satisficing.

Von Neumann and Morgenstern (1974) suggest another important facet of how respondents may satisfice, known as the lexicographic rule. Under this rule the respondent focuses only on the attribute most important to them, and selects the alternative that performs best on this attribute alone. If two options are equivalent on this attribute, they will select the next most important attribute to them and use a similar method to select the most desirable alternative based on this secondary attribute.

Another strategy used by respondents to make a decision is the conjunctive and disjunctive rules (Dawes 1964) which suggest the individual has a threshold level for all attributes. The conjunctive rule suggests an alternative surpassing all attribute thresholds will be an acceptable choice, while under the disjunctive rule, each alternative surpassing at least one threshold will be an acceptable choice.

Elimination by aspects (Tversky 1972) can be thought of as simultaneously using the lexicographic and conjunctive rules (Stüttgen et al 2012). The respondents will focus on the single most important attribute to them. All alternatives where the threshold is passed for this attribute make it through to the second round of the decision-making process for the respondent to make a final decision. Satisficing tends to occur more frequently in self-administered surveys as there is no interviewer present to keep the respondent attentive and motivated to respond to the best of their ability. Interviewers may also help to reduce task complexity by assisting the respondent to fully understand the question. Both functions help to reduce the respondent’s potential need to satisfice (Holbrook et al 2003). Holbrook et al suggest, amongst self-completion methodologies, online surveys tend to suffer from the effects of satisficing the least, as they tend to provide more sophisticated methods to reduce task complexity, keep the respondent engaged and motivate them to respond fully.

However, Lindhjem and Navrud (2011) analysed several studies that evaluated the extent of satisficing across different survey modes. They found mixed results and eventually concluded satisficing is no larger a problem in any one type of surveying mode over another. This suggests the extent of satisficing is more likely to be related to the survey design and respondents’ characteristics, rather than how the survey is delivered to the respondent.

To reduce the likelihood of respondents resorting to satisficing, Carlsson (2009) suggests task complexity can be minimised by providing visual aids which demonstrate scenarios in a way respondents can better understand, internalise and engage with.

Lindhjem and Navrud (2011) suggest another option to reduce the effect of satisficing is to use a pilot study to determine the extent to which respondents find survey questions complex, and simplify the survey accordingly. This can be done by asking respondents how complicated they found each question, whether they ignored any attributes in their decision process and why, and identifying questions consistently rated as excessively complicated, then adjusting these questions to reduce task complexity.

4.5.3 Effect of inertia and ‘non-trading’

Without significant external influence, individuals tend to be hesitant to change from what they are used to (Yanes and Ortuzar 2010). Hess and Rose (2009) describe a similar concept they refer to as ‘non-trading’. This is defined as an individual’s natural inclination towards the status quo rather than fully considering novel alternatives.

Carlson (2009) also observed respondents tend to favour default options most closely resembling the status quo, even when they have considered the status quo may no longer be the best option. Respondents justify this by stating the current system must have been put into place for a good reason, ‘so why should it be changed?’ This train of thought encourages non-trading as respondents fail to even consider novel alternatives they have been offered.

Studies which offer novel alternatives alongside the currently used scenario (perhaps as a point of reference) are particularly vulnerable to the effect of inertia or ‘non-trading’. Hess and Rose (2009) suggest techniques which can be incorporated into survey design to bypass this effect, eg asking respondents to rank alternatives in order from most to least favoured, rather than simply asking them to pick their single most favoured alternative. This allows for richer data collection in terms of subsequent preferences relative to one another, and is particularly useful to investigate secondary preferences when respondents tend to consistently favour the status quo option. However, as noted above, ranking adds considerably to complexity and is not how people usually make route decisions (section 4.2.1).

Hess and Rose (2009) also suggest the use of 'dual responses' to minimise the effect of non-trading. Initially the respondent is asked to rank or select their favoured option among two more novel alternatives (a forced choice). The respondent is then asked to make a second choice between their first choice from the initial round of questioning compared with the current status quo/reference option (a non-forced choice). It is suggested this technique minimises the effects of inertia as respondents are forced to fully consider new alternatives before making a final decision on their most preferred option.

4.5.4 Preference learning

Often respondents make inconsistent choices when completing SP studies. This does not necessarily indicate respondents are making decision errors or the survey design is flawed. Rather such inconsistency may be caused by respondents learning their preferences as they progress through the survey (Carlsson 2009; Day et al 2012). Groves et al (2004) suggest preferences are formed at the time the survey is completed, rather than simply waiting in the mind of the respondent to be collected. Preference learning is particularly prominent when respondents are unfamiliar or unexperienced in the subject area they are being asked about (Bateman et al 2008).

Brouwer et al (2010) undertook a SP survey which asked respondents how certain they were with the choice they had made after each question. They found respondents tended to become more certain about their answers as the survey went on, supporting the theory of preference learning. Respondents tended to feel significantly more confident about their choices at the end of the survey, compared with at the start. By way of a split sample test, Brouwer et al (2010) also found it was not necessary to ask respondents the degree of answer certainty for preference learning to occur. Many studies have documented the effect of preference learning and suggest techniques can be used to avoid it negatively impacting on data quality.

Braga and Starmer (2005) suggest preferences provided at the later stages of a SP survey tend to best reflect the respondents' normative preferences, shown by more coherent and consistent choices. Because respondents may have still been learning their preferences, questions should be repeated throughout the survey, with initial preferences discarded.

Carlsson et al (2012) tested this suggestion by completing a survey which comprised a single set of eight questions, repeated twice to create a complete survey of 16 questions. They found error variance was significantly higher in the first set of questions, compared with the second. Carlsson et al (2012) suggest this was not only due to preference learning, but institutional learning (learning how the survey itself works). Further, they found random variance decreases significantly after the first few choice sets. This suggests it does not take long for the process of preference learning to take place. As a result, Carlsson et al (2012) advised against including results from at least the first two questions; this would reduce error variance and increase the statistical power of the survey.

Alternatively, Carlsson (2009) suggested asking respondents to complete a 'warm up' survey prior to the main survey of interest. The warm up survey allows respondents to familiarise themselves with the structure of the survey so by the time they begin the main survey, their preferences will not be influenced by institutional learning.

It is recommended researchers test for stability of preference, as this indicates response validity (Carlsson 2009). Pilot testing may be utilised to measure preference consistency across equivalent questions to determine at what stage and to what extent the survey begins to elicit steady and accurate responses.

4.5.5 Questionnaire length and fatigue

Swait and Adamowicz (2001) observed, when respondents were asked to complete a similar task repeatedly, levels of random variation and inconsistency initially decreased before increasing again. These findings suggest survey structure, in terms of repetition and length, is an important factor which should be taken into account when creating a SP study. Surveys should be long enough so respondents are given the opportunity to learn and exhibit their actual preferences consistently, but not so long they tire and begin satisficing.

Carlsson and Matinsson (2006) investigated the effect of the number of questions or 'choice sets' respondents were asked to complete in a SP survey. They undertook a mail out self-completion survey in which the sample was split randomly. Half received a questionnaire with six questions, while the other half received a questionnaire with 12 questions. They found respondents who had received the longer questionnaire could handle answering many questions with only a small increase in non-response rate attributable to survey length. This suggests respondents are generally able to handle a significant number of questions as part of a SP questionnaire before a significant increase in non-response rates.

Galesic and Bosnjak (2009) investigated the effect of survey length on response quality for a web based survey. As expected, they found questions asked later in the survey were associated with a faster response time and more uniform responses than those questions asked at the beginning. This may be partially attributed to respondents' satisficing towards the end of an extended survey. However, this could also be attributed to respondents becoming more skilled at answering the survey in terms of both preference and institutional learning.

4.6 Stated preference survey design and modes

When designing a survey, it is crucial to investigate how the survey mode may affect results and outcomes of the study (Lindhjem and Navrud 2011). In particular, when designing a SP study it is important to consider how the survey mode may impact on the quality of the results produced due to the complexities associated with SP studies which we have previously discussed.

4.6.1 Online

This survey methodology involves sending out an invitation to take part in an online survey. It is a convenient and relatively cheap survey method which allows for large scale sampling to occur in a timely manner (Windle and Rolfe 2011).

Historically, internet surveys have been associated with representational issues in terms of both coverage and non-response error (Szolnoki and Hoffman 2012). This is particularly the case with samples drawn from an online panel (Duffy et al 2005). However, current literature is mixed, with some studies finding no difference in the accuracy of results when collected from perhaps the most scrutinised sample group; an opt-in internet panel, compared with random telephone or face-to-face interviews (Ansolabehere and Schaffner 2014). Windle and Rolfe (2011) suggest, with current high internet penetration and familiarity rates, online sampling is no more prone to errors than other frequently used survey methods. However, Szolnok and Hoffmann (2012) argue online panels are still likely to have major issues with representation. In a study conducted in Germany during 2012, Szolnok and Hoffmann found 4.7% of the population was currently on an internet panel, 20% of which actually responded to surveys. They concluded, as a result, only about 1% of the population is represented in online panels.

As a result, caution should still be taken to ensure online samples are representative of the wider population of interest. In the case an online sample is not thought to be sufficiently representative, data may be post-weighted according to population demographics. This is expected to align results more closely with what we would expect from a fully representative sample, although we note weighting will not solve a bias problem where the bias is not related to the weighting characteristics.

Bateman et al (2009) recommend the use of online surveying when undertaking a SP survey as it offers a more sophisticated and comprehensible medium by which survey contexts and alternatives may be presented. This may be particularly useful for SP surveys which are known to be relatively complex and often taxing on the respondent. Bateman (2009) suggests online surveying can be utilised to create a media rich, interactive and visual display which clearly demonstrates question contexts and associated alternatives; perhaps in the form of a virtual reality. He suggests such a technique is likely to boost respondent understanding, minimise cognitive load and enable respondents to make trade-offs in a way that mimics reality.

Raghothama and Meijer (2013) present a case for the use of gaming simulation in transportation analysis due to the interactive element it provides. They reference studies where gaming simulations have been used to successfully inform logistics research; which is said to have many links and parallels to transport research. Erath et al (2016) also undertook a virtual reality SP survey; however, they wanted to analyse potential cycle transport routes. They suggest an immersive virtual reality environment can be used as a supplementary tool, alongside other surveying methods, to understand how and why respondents make trade-offs between alternatives. Erath et al (2016) admit, due to the relatively new nature of this technique, potential limitations are yet to be fully discovered and understood.

4.6.2 Face-to-face interviews

Face-to-face interviews have long been considered the ‘gold standard’ of surveying, and have been said to work particularly well for SP studies (Arrow et al 1993). Although time consuming and relatively expensive, face-to-face surveys generally have low non-response rates and can produce high-quality data and in turn, robust results.

As with internet surveying, face-to-face interviews allow for the use of visual aids to assist respondents to understand and consider complex questions, albeit perhaps in a less sophisticated way than online surveying. For example, interviewers may be able to display diagrams or provide written information to respondents to help them answer survey questions. Further, the interviewers themselves may be able to answer any questions the respondent may have and provide additional explanations to ensure they have taken all information into account before answering the question (Lindhjem and Navrud 2011).

Having an interviewer present is known to motivate respondents to remain engaged with the survey so they not only are more likely to complete the survey, but complete it to a higher quality (Lindhjem and Navrud 2011). It has been suggested having an interviewer deliver a survey may reduce the likelihood of respondents resorting to satisficing.

Despite this, one of the key issues to be considered with face-to-face interviewing is the effect of interviewer presence on the respondents’ preferences due to social desirability bias. As distance between the respondent and interviewer decreases, the effect of social desirability bias increases (Lindhjem and Navrud 2011). Of all survey methods, face-to-face interviewing has perhaps the smallest distance between respondent and interviewer indicating the relatively large potential effect of social desirability bias. Comparatively, internet surveys are known to be significantly less influenced by social desirability bias compared with face-to-face (and phone) surveys (Kreuter et al 2008).

Surveys that include potentially sensitive questions tend to be particularly susceptible to social desirability bias (Kreuter et al 2008). Social desirability bias encourages respondents to overvalue socially desired goods, and undervalue less socially valued goods. In the context of a SP study, social desirability bias may encourage respondents to overvalue public goods (for example, safety) when an interviewer is present.

4.6.3 Telephone interviews/CATI

Telephone interviewing involves interviewers contacting and administering surveys via phone. It tends to be a mid-level cost option, as it is generally cheaper than face-to-face interviewing but not as cheap as online surveys (Arrow et al 1993). For this reason, telephone interviewing has historically had more use than face-to-face interviewing (Lindhjem and Navrud 2011).

Mitchell and Carson (1989) suggest telephone surveying is appropriate to use with SP surveys when respondents are already familiar or experienced with the contexts they are asked about. This is because respondents are reliant only on what they are verbally told by the interviewer to base their answers on. There is no opportunity to provide respondents with visual or written information when using phone interviews, unlike face-to-face or online surveys. Due to the relative complexity of SP surveys it is not unlikely for respondents to resort to satisficing, particularly if the respondents' pre-existing knowledge about what they are being asked is poor. For this reason, Windle and Rolfe (2011) advise against the use of telephone interviewing for complex and/or specialised CM surveys.

Groves et al (2004) found telephone interviewing to be the most susceptible survey method to social desirability bias, as there is less trust able to be established between interviewer and respondent. They suggest as interviewer trust increases, the effect of social desirability bias decreases (Groves et al 2004).

4.6.4 Mail

Mail surveying involves posting respondents a hard copy of a survey for self-completion and return. Again, this method has had more use than more involved survey methods (such as face-to-face interviewing) because of the relatively low cost (Lindhjem and Navrud 2011).

Similarly to phone interviewing, Mitchell and Carson (1989) suggest mail surveys are used only for SP surveys asking questions which respondents are expected to already be somewhat familiar with. Although mail surveys allow for some degree of visual imagery to aid in context visualisation, these techniques tend to be less sophisticated compared with what is possible using online surveys.

Although lack of interviewer presence may be a positive for reducing the effects of social desirability bias, it also means there is no one to assist the respondent in understanding and answering questions to the best of their ability (Lindhjem and Navrud 2011). As previously discussed, due to the complexity of SP studies, interviewers are a valuable resource which can be used to ensure the respondent fully understands what they are being asked to answer the question to the best of their ability. The absence of both an interviewer and sophisticated visual imagery associated with mail out surveys suggest such a methodology is not the best for complex SP studies.

4.7 Implications for a survey

The literature has several implications for survey design:

- Researchers have used route choices as the basis for a CM survey. This allows respondents to choose alternatives which differ with respect to attributes such as journey time, time reliability and safety (risk of crashes resulting in injury or fatality).
- There is no strong consensus on the number of alternatives to be included in a choice experiment. We have developed a design using two alternatives.

- A small number of attributes are easiest for participants to comprehend and respond to consistently. For this study, this needs to be balanced against the objective for the work to include as many attributes as possible in one survey so values can be derived for several transport outcomes.
- Labelled choices convey important information and should be avoided if the label has no significance or invites the participants to introduce non-modelled attributes.
- Wide attribute ranges provide better parameter estimates, but need to be realistic for participants to take them seriously.
- Heterogeneity and non-linearities can be modelled after the event; however, it may be beneficial to recognise their existence in creating the experimental design.
- The survey, while designed to be completed by a representative sample of the population, would usefully collect additional information so it could be used to identify the impacts on WTP of income, age of participant plus other information such as risk-appetite.

Other issues which need to be considered in the survey design include:

- How to deal with trip purpose and time of day (peak/off peak travel), because WTP may differ with these, and the associated differences in the opportunity cost of time.
- The road type, including the implications for risk level, ie whether to use road type as a way of making the risk level more realistic.
- What payment mechanism to use, eg a toll road, even though people in most parts of New Zealand will have no experience with such roads.
- What categories of injury to use, so the survey results can be grossed-up using national injury rates (see table 2.4 for New Zealand definitions).

A small-scale pilot study was run to test the design and the process. The pilot assessed whether robust valuations of non-market transport impacts including travel time and safety/risk and potentially other factors such as road quality, could be derived through CM methodologies applied to New Zealand based survey data.

Prior to designing the pilot, two rounds of initial testing were used to explore the implications of different survey designs. These alpha and beta tests are described in chapter 5.

5 Alpha and beta testing

5.1 Alpha test

5.1.1 Description of the test

An alpha test was undertaken to better understand how people conceptualise, evaluate and trade-off the non-market transport attributes under test. It provided an initial evaluation of questionnaire performance via a series of face-to-face, semi-structured interviews with a small number of selected participants.

In evaluating a question's performance, cognitive testing examines the question-response process. It can be conceptualised by four stages: 1) comprehension, 2) retrieval, 3) judgement and 4) response. It considers the degree of difficulty respondents experience as they formulate an accurate response to the question.

The interviews were designed to elicit respondents' thought processes when answering the tested questions, specifically, how they understood a question and how they arrived at their answer. Data from cognitive interviews is qualitative, and analysis of these interviews can indicate the sources of potential response error as well as various interpretations of the question.

Researchers sat alongside participants as they completed the draft questionnaire online, or on a laptop, to replicate the expected final survey implementation approach. They observed how respondents interacted with the questionnaire, and regularly probed to get feedback on comprehension, ease of responding and the thought process for each question.

5.1.2 Presentation of choice task

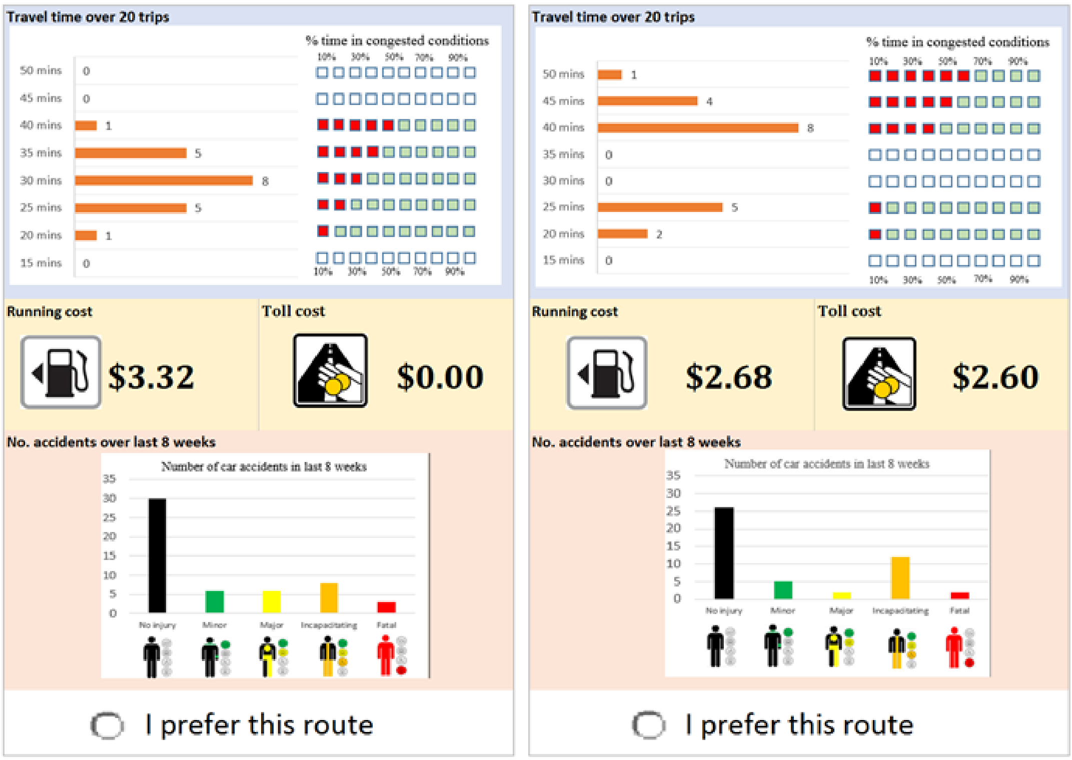
In developing the survey for testing, in addition to the lessons from the literature review, we based our design on the approach used in a recent Australian study, which include the following attributes (John Rose, pers comm):

- travel time in free flow and congested conditions, and the distribution of travel time between these
- costs, including running costs and toll charges
- crash profile described by average number of crashes by level of severity in recent weeks.

An example choice task formulation from this study is shown in figure 5.1. The numbers of crashes are raised to levels well above actual risks for any roads, but enable a better analysis of the trade-off between time, reliability, cost and risk of injury or fatality.[28] The crash risk information assumes respondents regard the number of crashes in the past eight weeks as a measure of their personal risk of being injured or having a fatal crash.

[28] This assumes the value of marginal changes in risks is independent of the initial levels.

Figure 5.1 Single trip question



Source: John Rose (pers comm)

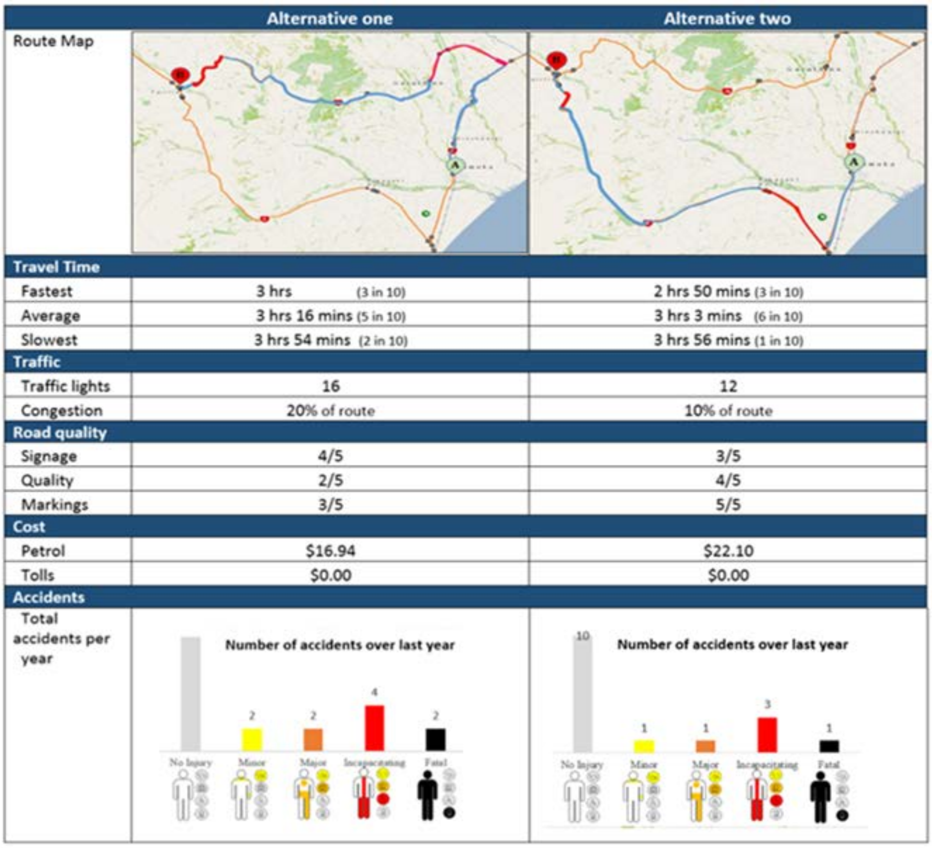
Building on the figure 5.1 example, draft questionnaire design options were developed, to allow further assessment of:

- Real or hypothetical journey – a hypothetical journey was used, but this was anchored to a real trip type, eg a regular commute or a regular trip to the next main centre. Respondents were asked to provide information on a recent trip, and the example trips used in the survey were then some multiple of this, eg 10% longer or shorter.
- Number of choice tasks – versions with five and eight choice tasks were used.
- Number of attributes. Those examined were:
  - travel time – average journey length
  - time reliability – fastest and slowest times
  - traffic:
    - number of traffic lights
    - % of route in congestion
  - road condition and quality
    - signage (score out of 5)
    - road quality (score out of 5)
    - markings (score out of 5)
  - cost – fuel and tolls
  - crashes – number of crashes by injury type (no injury, minor, major, fatal)

- Presentation of attributes – half the participants were given a mock GPS layout (figure 5.2) and half were given a simplified grid (figure 5.3). At the end, they were also shown a version with no map (figure 5.4).

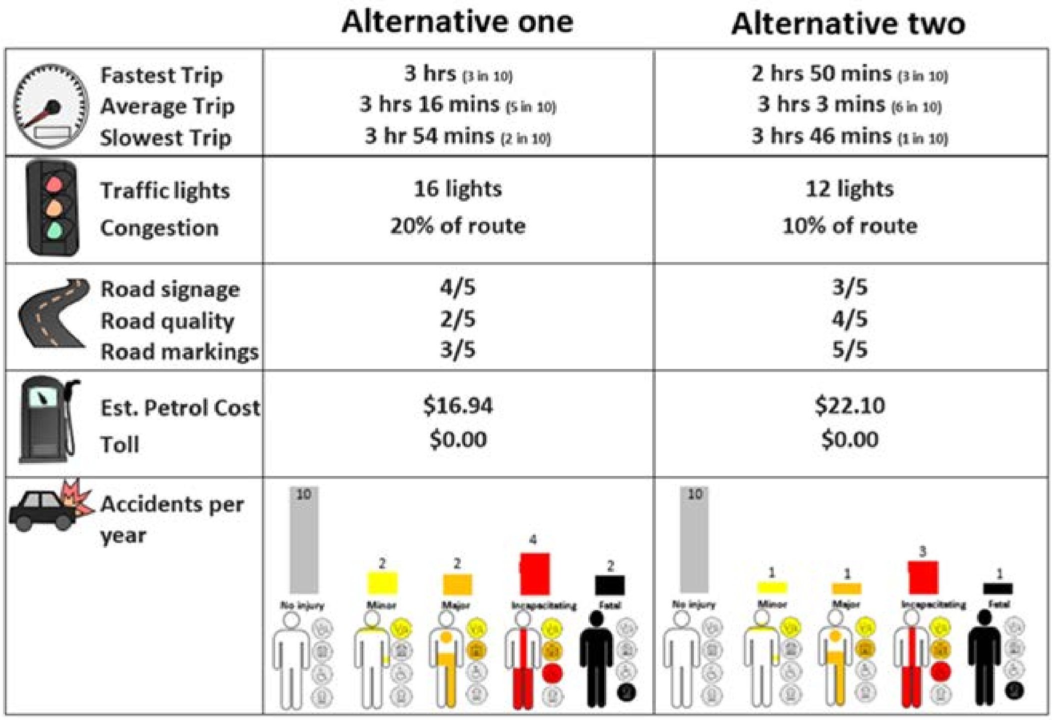
Figure 5.2 Screenshot of alpha test GPS layout

Figure 5.3 Map/grid layout



	Alternative one	Alternative two
Route Map		
Travel Time		
Fastest	3 hrs (3 in 10)	2 hrs 50 mins (3 in 10)
Average	3 hrs 16 mins (5 in 10)	3 hrs 3 mins (6 in 10)
Slowest	3 hrs 54 mins (2 in 10)	3 hrs 56 mins (1 in 10)
Traffic		
Traffic lights	16	12
Congestion	20% of route	10% of route
Road quality		
Signage	4/5	3/5
Quality	2/5	4/5
Markings	3/5	5/5
Cost		
Petrol	$16.94	$22.10
Tolls	$0.00	$0.00
Accidents		
Total accidents per year	Number of accidents over last year	Number of accidents over last year

Figure 5.4 Grid layout with images but no maps



One of the key differences from the Australian survey (figure 5.1) was the use of a simpler version of the journey time reliability presentation. In the Australian survey, a distribution of journey times was used across eight durations (in five-minute categories from 15 to 50 minutes). We have used a simpler version which emphasises the average travel time, while still providing some information on reliability (variability). Our assumption in making this simplification is the focus of the respondent is on expected travel time and they will be using the distribution of travel times to calculate an expected travel time. However, it is not obvious all respondents will correctly (or equally) estimate the expected travel time from the information provided. The approach adopted is to specify the expected travel time as the average time, while still providing some estimate of the variance.

The survey assumes people have perfect information about costs, but are working with probabilities relating to journey time and crash risks, from which they have to interpret their safety risk (and note personal risk is of interest here).[29] The issues to be considered in the survey design phase are the extent to which this type of formulation is too complicated for people to understand and make reliable decisions from. To assess this, in the initial round of qualitative interviews, as well as reviewing how easy respondents found the process and what they understood the information to mean, we discussed how people formed their decisions, what information they relied most on and the extent to which (if at all) they simplified the information presented.

Supplementary questions asked include those requesting basic information about the respondent (age, income etc) and whether they have had a car crash in recent years and the nature of the injuries, or if they know someone who has experienced a major crash; this experience might provide them with more information on the relative costs of injuries and fatalities.

---

[29] See discussion in Viscusi (2005)

Testing of surveys was undertaken to develop these ideas further and focus particularly on the balance between:

- a survey to obtain the maximum amount of usable data, and
- simplicity and understandability of the survey for participants.

After completing the questionnaire, participants were asked a series of questions about the experience, including identification of questions the respondent found difficult to understand or answer, and assessment of how the questions were answered, what factors were considered and how response decisions were made. The researcher referred back to key questions/formats and reviewed these in detail with the participant.

We conducted 15 interviews face-to-face with road users, using qualitative researchers experienced in both in-depth interviews and in questionnaire design and testing.

Participants were selected to provide a mix of people including socio-economic levels, ethnicities as well as driver types (commuters, non-commuters etc) and included urban, provincial and rural dwellers. We interviewed participants in Auckland (eight), Rotorua (four) and rural Waikato/BoP (three). Interviews took around one hour at people's homes, or a central location and participants were paid an incentive to thank them for their time.

Interviews comprised two main elements:

- Discussion of journey/route attributes, including: the factors considered when making journey choices where alternative options exist; discussion of proposed attributes, how these are conceptualised, valued and traded off and how well these align with real-world considerations, including whether any journey attributes are deemed irrelevant or missing.
- Testing and review of draft questionnaire approaches, focused on the choice set exercise element and comprising accompanied completion of questions. This assessed comprehension, retrieval, judgement and response and the degree of difficulty respondents experience as they formulate their responses and choices.

Researchers assessed these factors through observation and discussion both during and after question completion. The interviews were designed to elicit respondents' thought processes when answering the tested questions, specifically, how they understood a question and how they arrived at their answer.

5.1.3 Results

The main results of the alpha test are summarised in table 5.1 and discussed below.

Table 5.1 Alpha test results

Component	Results
Presentation of choice tasks	Many found the maps useful and about half the respondents said they had referred to them in making their decision, eg how direct the route was. This was problematic because the map was becoming a source of information in an uncontrolled way (see discussion below).
Number of choice tasks	Initially five choice tasks were presented. This was increased to eight because early participants indicated they could undertake more. Subsequently some indicated they were losing concentration towards the end. Respondents became more certain in their responses after the first or second choice tasks.
Number of attributes	No participants said there were too many attributes to consider, but some noted there was a lot of information and some attributes were ignored, either because they did not think them important or because of information overload.

Component	Results
Realism	Some participants started to exclude factors they did not normally consider in route choices. And some noted they usually just rely on a GPS/Google maps. However, the task was to encourage respondents to consider all relevant information, regardless of whether they normally did.
Journey time	Journey time was the attribute most often regarded as most important in the choice decision, but participants would disregard small time differences. Most ignored the fastest and slowest trip information and focused on average (expected) travel time.
Congestion	Respondents understood congestion and regarded it as important in their decision. Some relied on the congestion information in the maps rather than the stated percentages; this was problematic because we were not controlling for this in the analysis of results.
Fuel and other cost	Many participants did not consider cost in their decisions, partly as it was regarded too small to influence the choice. Toll costs were included but many participants reacted to the idea of paying tolls, including those thinking the survey might have been about introduction of tolls and rejected tolled routes in protest.
Crash and injury risk	About half the participants did not consider crash risks in making route choices. This partly appeared to be because of the way the information was presented; the number of injuries and fatalities did not translate into a risk for them.
Traffic lights	There was a mixed reaction to traffic lights. About half the participants did not consider the traffic lights, some favoured them and others avoided them. There was potential for interaction with the congestion attribute which was not treated the same by all participants.
Road markings and signage	These features received lowest priority in route decisions. In addition, the scale used (1 to 5) was not easy to interpret.
Trip purpose	This was added part way through. This made the game less abstract and helped participants to make decisions.
Vehicle type	Participants commented that the type of vehicle they used made a difference to their attitude to safety.

5.1.3.1 Game layout

Participants were given one of two layout alternatives to complete the survey. At the end of the survey they were asked to review three alternative screen layouts for the paired choice games and to comment on the strengths and weaknesses of each and their preferred layout. There was a mixed reaction.

- About half of participants liked having a map, crash and other images and used the maps and crash images in some way to influence their decision. Of the other half who did not rely on images, none said they disliked them or wanted them to be removed.
- While some participants said they preferred the GPS layout, none strongly preferred it or said they disliked the other layouts. However, some disliked the GPS layout mainly because they found it more cluttered and felt the information was easier to find and compare across a row when presented in more of a grid format.
- Most participants said they favoured more simple layouts with fewer colours and four of the participants suggested the GPS layout was at first overwhelming compared with the other layouts.
- The GPS screen layout showed the crash/injury graphic as a separate screen below the main information screen. Two participants said they had already decided on their preferred journey choice before they scrolled down to the crash information.

There is a need to respond to the fact participants used information presented in the maps rather than the numbers presented. This is an example of a labelled experiment. Simplifying the maps to reduce this effect was trialled in the beta test.

5.1.3.2 Number of attributes

Few people indicated they had considered all attributes as they progressed through the games. Fuel cost (for single journeys depicted), road signage, quality and road markings were most often identified as the least important factors to participants when deciding which alternative to select. Arguably, statistical analysis of a large data set (from a wider survey) would demonstrate that many people have considered these factors, but simplifying the choice games and providing respondents with fewer attributes to evaluate is likely to improve the statistical significance of the parameter estimates, particularly in a smaller survey.

Although cost was identified as a factor which people in this test version did not think important, it was necessary to include it because we will always need to understand the way people trade-off cost for other attributes of a journey.

In the pilot test (see below), it was made clear we wanted participants to evaluate all attributes they are presented with, even if they were attributes they would not, or could not, consider in real life. To assist with this, the number of attributes was reduced.

5.1.3.3 Variation within each attribute

Most participants indicated the differences between the choice pairs offered tended to be quite small with attribute levels quite similar in each pair and that attributes did not vary much across the games.

A significant difference between values within each attribute may help participants to make decisions and avoid people ignoring attributes entirely (once they had seen little variance). However, care must be taken not to use unrealistic values.

Participants mentioned this in reference to fuel costs, travel times on shorter trips and crash numbers.

Increasing the differences among attributes to the extent possible, while still covering the full range of possible values for an attribute across all participants would address this issue.

5.1.3.4 Strength of preference

We noted some participants felt several of the sets were very hard to choose between as they were very similar, and remarked they had chosen an option only because we asked them to and in reality, would have been happy to take either route.

A strength of preference question after each choice (or as the choice device) could improve the respondent experience. It may also help to indicate preference learning (Brouwer et al 2010) as participants learn what their priorities are as they progress through the games. A strength of preference was added to the beta test questionnaire (see figure 5.5 below).

5.1.3.5 Journey time

Average journey time was regarded as very important to the choices made, but travel time reliability information was often ignored. As noted above, our assumption is people make decisions on the basis of expected travel time and the inclusion of an average time reinforces this. However, to the extent uncertainty is an issue or time reliability is an attribute of interest, this approach does not necessarily provide a good understanding of the value ascribed to it.

5.1.3.6 Safety risk

Effectively communicating injury risk by severity is challenging, because of the complexity of the information and a likely disconnect between the data presented and how people conceptualise and deal with injury risk in practice. It appears safety risk is not something people generally consider, so they do not have a relevant frame of reference for the (realistic historic) data provided.

Alternative formulations of injury and fatality risk were used in the beta test to examine this further.

5.1.3.7 Fuel cost

Several participants stated they did not take account of costs (especially for short journeys) or they objected to the mechanism (tolls). For this study, cost had to be one of the parameters included and it was important it was taken into account by participants, because we wished to place a value on the other parameters in monetary terms. The survey would be more likely to yield useful results if it focused on the types of journeys for which participants were most likely to consider costs, eg longer trips. However, this needed to be balanced against the interest in exploring non-linear relationships, eg if values differed with trip length.

5.1.3.8 Survey interview method

Of the 15 participants, three said they found the survey interesting and relevant and if they had been sent an online survey with no incentive, they would still be likely to have taken part in the research. However, the remaining participants said they were either not sure if they would take part in the research independently, or they would not.

Most participants also felt they had put more thought into the games than they would if they were completing the survey alone and without the knowledge required to answer questions about their decision-making process 'I knew you'd ask me why I said what I said… I thought way more about it but don't know if I'd do it on my own. Hard to say how much I'd think into everything on my own.'

Holbrook et al (2003) talk about satisficing being more common in self-administered surveys as there is no interviewer present to keep the respondent attentive and motivated to respond to the best of their ability. The alpha test participants acknowledged they would be less likely to give the survey their full attention if self-completing. This desire to expedite the experience, coupled with the complexity of the survey and the need to explain both the process and some of the attributes if they are to be properly understood, suggests using an accompanied survey method (face-to-face interview, virtual accompaniment, or a telephone follow up) should be considered. Alternatively, if the survey were to be conducted using an online method, reducing the complexity of the information, starting with the number of attributes shown, should be considered.

5.1.3.9 Trip purpose and distance

Trip purpose was added part way through the testing and made a positive difference to the ease of people's decision making. Trip distance was varied with trip purpose also. Participants clearly expressed differences by trip purpose, for example, journey time can be less important when on more leisurely or discretionary trips, say visiting a friend, compared with a work trip, or commute.

Trip purpose helped make route choice games less abstract, with the choices easier to make where they could be related to real-life mindsets when embarking on different types of trip.

A few participants commented that even more information and context would have helped with their decision making, as other factors can be influential, for example, whether the trip is a first-time trip (if visiting a friend whether you know where the friend lives, or have visited there before), or the time of day of the trip (whether in peak traffic, inter peak/evening, or a weekend).

5.1.4 Suggested changes

The following key conclusions were drawn from the alpha test.

- Because people are using information in the map it is difficult to account for this in the analysis. If maps are to be used a much simpler version is required.
- Seven or eight choice tasks are likely to be a suitable number.
- It would be useful to reduce the number of attributes to ensure the attributes of main interest are considered, particularly cost and injuries/fatalities.
- The variability or reliability of journey time needs to be presented in a way that is readily understood.
- Providing information on trip purpose helped people to make decisions and we recommend its continued use.
- Costs need to be included at a level at which people will take them into account.

5.2 Beta test

5.2.1 Presentation of choice task

A beta test was used to examine the effectiveness of changes made following the alpha test. The aim of beta testing was to eliminate as many faults/bugs as possible before the pilot test. Since it is plausible this type of survey is conducted face-to-face with an interviewer, or via self-completion online, we tested the survey using both these methods.

A larger number of participants than in the alpha test was used: 40 interviews were conducted, 15 face-to-face and 25 by sending recruited people a link to the online questionnaire. The questionnaire was implemented as it would be in the standard survey process; post-completion questioning was used to assess ease of completion and comprehension of key concepts and questions. These were asked by the interviewer in the face-to-face survey and via a post completion phone call from a researcher in the online survey.

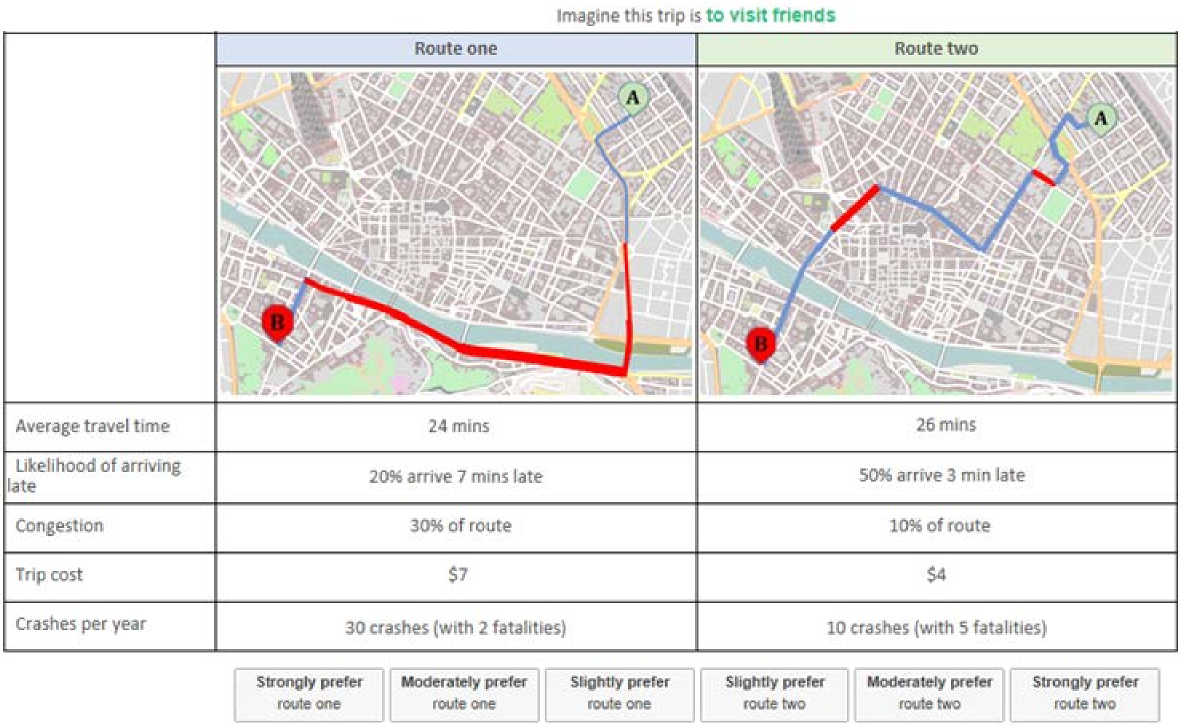
Because the GPS question format in the alpha test was too complex, an alternative format was used to convey choices in the beta test (figure 5.5). Respondents were asked to choose between two competing routes. The attributes of the routes were presented underneath the map.

The attributes used in the beta test were:

- travel time – average trip duration
- trip time reliability/variability – likelihood of arriving late (two attributes: probability and lateness)
- congestion levels – % of trip length spent moving slowly and/or stopped
- costs – explained as including fuel, other running costs and possibly tolls
- injury-risk based on number of injuries and fatalities (two attributes).

We also used a version with numbers of crashes over the previous 12 months by injury level (no injury, minor injuries, major injuries) and the number of fatalities.

Figure 5.5 Beta test: GPS presentation of proposed choice tasks



In the beta test, representative trips were generated by asking the participant to describe a recent trip. Priors were based on assumptions used in the Australian study and values from the EEM. These were used to generate an experimental design based on the attribute levels in table 5.2.

Table 5.2 Range of parameter values

Attribute	Value
Travel time (anchored on value provided)	5%, 10%, 25% (difference from representative trip value)
Likelihood of being late	10%, 20%, 30%, 40%, 50%
Lateness	10%, 20%, 30%, 40%, 50%
% of route congested	10%, 20%, 30%, 40%, 50%
Trip cost	Various levels based on representative trips, cost per km and a hypothetical toll
Total crashes	5,10,20,30,40
Fatal crashes	1,2,3,4,5

5.2.2 Results

The results of the beta test are summarised in table 5.3 and discussed below.

Overall, participants found the beta test to be easier to understand, and they paid more attention to attributes than in the alpha test. Nevertheless, there were a number of comments that suggested improvements were still required.

Table 5.3 Beta test results

Component	Results
Presentation of choice tasks	Fewer comprehension issues were observed compared with the alpha test, due to a simplified layout and the reduced number of attributes for consideration. Some alpha test participants were re-contacted and given the beta version, and all commented it was easier to follow and understand; some of the returning participants noted they had thought about the process and their choices in the intervening time and approached the task slightly differently. Participants used the maps, often using information from them not covered in the attributes provided, eg route directness, number of intersections or how scenic or hilly they thought the routes may be. This was a problem not eliminated by using the simpler map. Face-to-face participants were observed skipping instructions and online participants spent less time reading instruction screens, suggesting not all information was read or understood and some respondents would inevitably take shortcuts. Face-to-face participants overall paid more attention to choices and information provided, with four of the fifteen face-to-face participants asking questions about anything they were unsure about. Also, more comprehension issues were picked up face-to-face that were not evident in the online approach.
Number of choice tasks	Eight choice sets were used. No beta test participants said there were too many choice sets.
Number of attributes	All participants found the number of attributes to be manageable. Six participants said they had 'ignored' some attributes altogether, but because they thought them less important, rather than because they were overwhelmed by the number to consider.
Realism	Four participants felt numbers such as congestion were not representative, or lateness did not apply to them as they lived somewhere with no congestion, or always left early, so would not be late. These people said their answer would be different if they had answered using the level provided rather than answering how they had done something.
Journey time	Participants made use of the average journey time information as the most important attribute in the trip choice.
Trip purpose	Trip purpose information was added to better enable participants to envisage the trip. Three options were included: two types of short local trip (regular commute and irregular local trip) and a long trip.
Time reliability	Reliability (or variability) of trip time, measured as lateness, was now more visible than previously, and participants mostly understood it, but very few indicated they considered it important to their route choice decision.
Congestion	Congestion was indicated as the second most important attribute when the process was discussed with alpha test participants; however, this appeared to decrease in the beta test, with more participants now considering cost and crashes. This may be an effect of greater explanation and simplification of the choice sets.
Fuel and other cost	Costs were increased in size and more participants took them into account.
Crash and injury risk	More participants considered crashes than in the alpha test, but for some this was because they perceived crashes would cause delays, rather than the risk of injury or death. Some people found it hard to internalise the risk of crashes. Some participants said they needed more information about crash frequency to help understand whether the number of crashes and fatalities presented was high or low compared with the average. Fewer participants said they ignored or dismissed this attribute than did in the alpha test. More related crashes to road conditions or features outside their control. This validated the greater explanation offered in the beta version.

5.2.2.1 Cognitive difficulty

Participants coped well with the survey. Their general impressions were it had an appropriate number of attributes, the information was presented clearly, and they could understand the task. There was some evidence of learning from the first-choice situations that framed the way the participants processed later choices. There was a low level of engagement by some participants, who sought to streamline the process. However, this does not necessarily mean the survey was too difficult should they have wished to engage.

5.2.2.2 Attribute non-attendance

Many participants stated some attributes were ignored or 'had little impact', eg likelihood of arriving late (n=37), congestion (n=31), crashes (n=26), cost (n=22), travel time (n=13). It should be noted 'had little impact' can be an artefact of the level (or range of levels) of the attribute, or of its salience. It appears that perceived very small differences in costs were of little impact, rather than cost being a non-salient attribute.

5.2.2.3 Imposed attributes

Maps provided unintended cues for respondents, who used information from the maps to develop their own attributes associated with the route alternatives. Further, participants imposed their own values (eg 'there is no congestion where I live', or 'I'm a safe driver so I won't be involved in a crash'). Some participants (5 of 14 who said crashes were important to them) viewed disutility from crashes as potential delay to them completing their own journey, rather than the risk they would be involved in a crash.

5.2.2.4 Attribute information

Some participants sought more information on crash frequency norms to enable them to assess risks. However, similar to the alpha test, those provided with information about the types of injuries or crash outcomes found they did not distinguish each crash type when assessing this attribute.

5.2.3 Responses

A number of conclusions were drawn from the beta test on how the survey could be further improved to increase understanding and engagement.

5.2.3.1 Maps

The maps help the respondents to better picture the choice sets but they contain significant amounts of information which might be used in making decisions. The use of maps makes this a labelled experiment. Survey participants can evaluate route attributes which they surmise from the map in making their choices, confounding model estimates. This other information processed by survey participants is unknown to the analyst, so cannot be modelled. Maps invite participants to substitute their own attribute values, based on their experiences with roads which look similar, eg they may not believe the travel time proposed by the analyst and may substitute their own 'superior' estimates. The analyst will be unaware this has occurred.

The options which need to be considered include dropping the map or making it very simple so it does not influence the choice. In practice, the only way to be certain we have controlled all variables being used (or have to the extent possible), is to not use the map.

5.2.3.2 Injuries and fatalities

A problem with the existing approach is people are not being encouraged to internalise the risk of injury and death, eg they are focusing on the problem of crashes causing delay. This is partly because the question format presents participants with information on the number of crashes rather than as their risk of being involved in a crash and the potential consequences. A revised format is required for the pilot test.

One approach would be to simply state the risk to them, eg ‘using route A, the risk of a crash in which you will be injured is x% and the risk of your having a fatal crash is y%’. The difficulty with this approach is, if the numbers are to reflect reality, the risk levels would be very small, eg the number of road fatalities was 328 in 2016 with total VKT of over 45 billion; this represents an average of 0.72 deaths per 100 million km travelled. For light vehicles only, the risk to occupants is 5.4 deaths per 100 million km travelled. The risk for any individual trip is thus tiny; using the average for light vehicles, the risk for a 100 km trip would be 0.00000054. So even if one route was half as risky, the number would still be very small.

This injury/fatality question is a crucial issue for the pilot.

5.3 Recommended changes for the pilot test

The beta test suggested a number of changes for the pilot test.

1. The presentation of the choice tasks using maps is best dropped to avoid the problem of respondents using the information in the maps as inputs to the route decision. The suggested approach is to use some pictures to illustrate the attribute, but in a way which does not convey any information about the attribute level.
2. Use a broader range of costs. This may present some difficulties because in the absence of toll costs, differences between routes are effectively because of differences in vehicle operating costs. They, in turn will be highly correlated with distance. However, if distance is the same for both routes, then there is no basis for cost differences.
3. Provide information on injury and fatality frequencies relative to the average so respondents can better understand relative risk.

6 Pilot test

6.1 Purpose

The pilot test aimed to trial the survey in a form as close as possible to one that might be used in a full survey. It aimed to identify whether the survey might produce results which could be used to identify a VoSL and values for other attributes included in the survey. The survey is reproduced in appendix C.

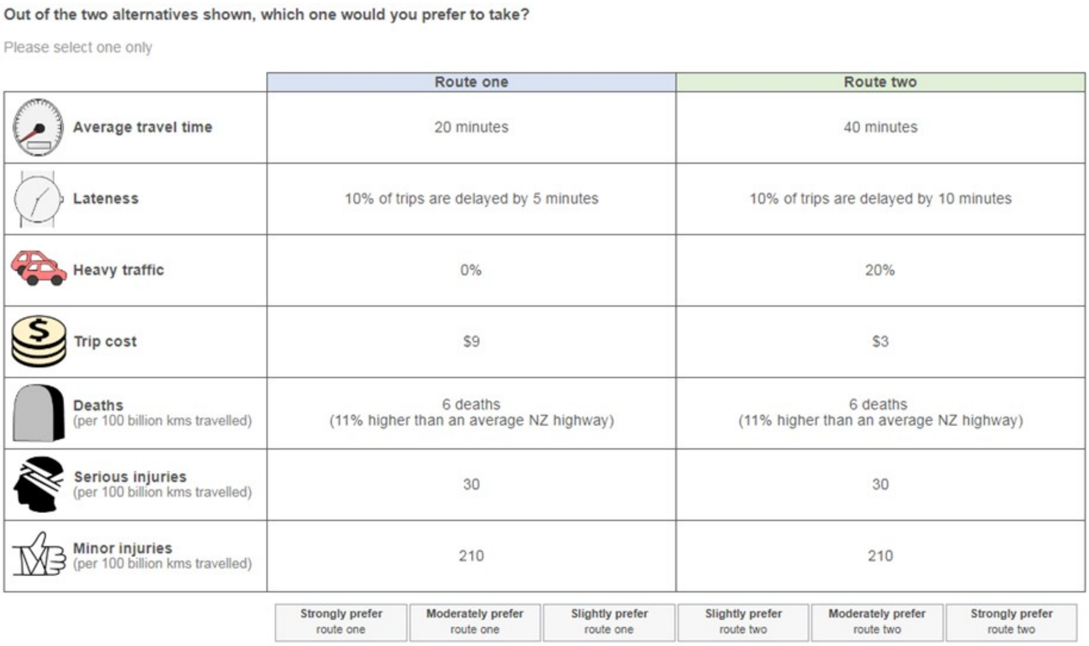
The pilot was also used to test survey approaches. Previous rounds of testing have involved online surveys and accompanied face-to-face surveys. The pilot aimed to assess differences between face-to-face and online interviewing, including differences in concentration, attentiveness, interview length, responses as well as the viability or ease of delivering the survey through an interviewer, compared with a self-completion survey.

After the alpha and beta tests, changes were made to simplify and decrease the number of attributes in each choice set, attribute levels and layout in the choice questions. Wording changes and further explanation were added to better define the attributes, particularly around the risk of injuries and fatalities.

6.2 Presentation of choice tasks

The choice task presentation used in the pilot test is shown in figure 6.1.

Figure 6.1 Pilot test choice task – example



Out of the two alternatives shown, which one would you prefer to take?

Please select one only

	Route one	Route two
Average travel time	20 minutes	40 minutes
Lateness	10% of trips are delayed by 5 minutes	10% of trips are delayed by 10 minutes
Heavy traffic	0%	20%
Trip cost	$9	$3
Deaths (per 100 billion kms travelled)	6 deaths (11% higher than an average NZ highway)	6 deaths (11% higher than an average NZ highway)
Serious injuries (per 100 billion kms travelled)	30	30
Minor injuries (per 100 billion kms travelled)	210	210

Strongly prefer route one	Moderately prefer route one	Slightly prefer route one	Slightly prefer route two	Moderately prefer route two	Strongly prefer route two

A number of changes have been made from the beta test:

- The map has been removed but graphics are retained in the form of the pictures to accompany the attribute titles. This is because of the extent to which people were using the map information as an input to their decision.

- The lateness attribute has a standardised percentage of trips which are late, but the extent of lateness (minutes) varies. This contrasts with the beta test where both percentage and minutes varied, which might have been two points on the same distribution.
- Heavy traffic (congestion) is presented in minutes rather than as a percentage of the trip.
- Crash impacts are presented in terms of the number of fatalities, major and minor injuries on the road per billion kilometres and the comparison in percentage terms to the New Zealand average.
- Respondents are asked to choose one option and to state the degree of preference.

6.3 Attribute numbers and levels

Deciding the number of attributes involves a trade-off between:

- the desire to include as many attributes as possible in the survey so values for several attributes can be derived
- the cognitive ability of survey respondents to weigh up multiple attributes.

The testing has shown the number included in the beta test was manageable. Some attributes were ignored or given little weight, but this appeared to be based on perceptions they were unimportant rather than because of cognitive difficulties.

The key attributes of the survey are those relating to the risk of being injured or killed in motor vehicle crashes, particularly fatalities so a VSL could be estimated, and travel time. Other attributes are included because they are of potential policy interest (eg congestion).

The problem of balancing complexity with realism is discussed widely in the literature, eg Cherchi and Hensher (2015) note:

> *In order to ensure realism and reduce potential hypothetical bias, analysts may need to build rather complex survey tasks which respondents are asked to process in a short time, potentially exerting high burden and risking damaging the quality of response. Individuals have limitations in their capacity to process information, and are not always willing to invest the required degree of effort in evaluating alternatives. When presented with a complex task, it is then likely that they show disengagement, adopting simplifying strategies to reduce the mental effort required to solve the problem. On the other hand, simplifying the survey tasks to reduce the cognitive burden for respondents is also risky. Simplified survey tasks are often too simplistic and they can be seemingly perceived as unrealistic by the respondents, leading to problems with respondents' engagement, or respondents choosing based on other attributes not included in the design. The risk is that a design with too little alternatives or attributes or levels might produce more errors than more 'complex' designs.*

The survey as illustrated above enables the identification of the value of time, a value of reliability, of avoiding congestion and values to avoid injury or death from road crashes.

6.3.1 Travel time and reliability

Travel time reliability represents a complex attribute that can be measured in multiple ways. For example, reliability can be thought of as the probability of arriving on time; however, research has shown arriving later than the preferred time is valued differently from arriving earlier, which is not always viewed positively. An attribute that only gives the probability of arriving on time will provide only partial data on the issue. Within the transportation literature, reliability is more commonly thought of as a distribution of travel times representing the probability of observing a particular travel time outcome (as in figure 6.1). Using this approach, travel time and travel time reliability are represented by the same attribute, with travel time typically being defined as the expected travel time given the travel time distribution, and reliability being the variance of the travel time distribution. However, it is not obvious all respondents will correctly (or equally) estimate the expected travel time from the information provided, and it is not clear whether they are responding to expected travel time or to the variance. The approach used for the pilot test, as with the alpha and beta tests, has simplified the choice task by specifying the expected travel time (the average time), while still providing some estimate of the variance.

6.3.2 Safety

People have difficulty processing very small probabilities, which has led to adoption of the approach in figure 6.1 (the Australian example) that provides survey participants with information on the total number of crashes on each route in a specified period. This approach has been employed previously (Antoniou 2014; Hensher et al 2009; Niroomand and Jenkins 2016; Rizzi and Ortuzar 2003; Rouwendal et al 2010). However, this specification of the crash attributes does not permit recovery of VoSL without making some extremely strong and untenable assumptions.

The studies cited above that have used total crashes have adjusted estimated per trip values to mean VoSLs through ex-post scaling by national fatality rates and trip distance. However, those fatality rates, which effectively convert trip fatalities to exposure levels, are unknown to the individual at the time they make their route choice decision, yet are presumed to be implicit in it. One way to overcome this is to provide participants with information on the probability they will be killed in a crash on the route. This probability will depend on the expected annual fatalities on the route, and the expected annual number of vehicles travelling the route. In reality, the probability of the individual being killed will be extremely small, and differences in fatality probabilities between routes will be even smaller, encouraging participants to ignore the attribute. A partial solution is to equalise traffic volumes on routes, so differences in annual fatality rates are proportional to relative risk. Adjustment still needs to be made for exposure per vehicle or passenger per journey. Provision of information on average traffic volumes does that. An alternative, adopted here, is to report exposure rates in terms of vehicle distances travelled and report either route lengths or inform participants that routes are of equal length.

The approach used in figure 6.1 anchors on the average national fatality rates for the road type and provides information on relative risks compared to the average. The relative fatality and injury rates on the routes allow participants to evaluate both relative risk.

6.3.3 Cost

The trip cost attribute is critical for derivation of monetary values because the cost attribute coefficient is the divisor in an equation used to estimate WTP. Hence, it is critical the cost coefficient is statistically significant and has a relatively narrow confidence interval. This is best achieved with a broad range of attribute values.

In many studies, changes in attribute levels are motivated by externally set fares (eg public transport) or externally set vehicle toll charges. Neither of these applies in the car route choice context proposed here; tolls have not been used because they are uncommon in New Zealand and because of the reaction seen in the alpha test in which some respondents avoided routes with tolls because they objected to them on principle and/or thought this is what the survey was about.

Consequently, the only justification for change in costs is differences in vehicle operating expenses between routes. This can present some statistical and/or conceptual challenges. On the statistical side, extremely high correlations between travel time, travel costs and distance are problematic, particularly in RP analysis. Multicollinearity means the effects of these three variables are confounded. In the hypothetical context confounding can, in theory, be overcome through appropriate experimental design in which these attributes vary independently of each other, allowing identification of their separate effects. However, that strategy can be a practical challenge if participants do not find the design combinations to be credible. For example, a very long trip with a very short travel time might seem quite unrealistic – and the perception of reasonableness might be influenced by other attributes, such as time spent in congested conditions. One solution is to fix one of the attributes and vary the attribute of interest (eg fix distance, allow time to vary if the interest is in valuing travel time). However, to value either time or distance, the cost attribute must vary. This raises the issue of credibility of changes in cost when distance is fixed.

The approach adopted attempts to include significant differences in the cost attribute while not making the cost differences unrealistic to respondents.

6.4 Utility functions

The utility functions which will apply to the attributes being measured are discussed here. Linear utility functions have the specification:

$$U_i = ASC_i + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + ... + \beta_{ni}X_{ni} \qquad \text{(Equation 6.1)}$$

Where there are n attributes, X_{ji} is the level of attribute j in choice alternative i, β_{ji} is the marginal utility of attribute j in choice alternative i, and the 'alternative specific constant' ASC_i is a constant associated with alternative i. There is a utility function specified for each alternative, with one ASC_i set to zero for identification purposes. With labelled choices, some attributes may be absent from some alternatives and the ASCs measure differences in utility that do not arise from the measured attributes. In situations with labelled choices it can make sense for the β_{ji} to vary across alternatives, permitting utilities to differ across alternatives (eg the value of travel time may be different in a train from in a car).

Generic utility function specifications are somewhat simpler. The utility of attributes does not vary by alternatives. The specification is:

$$U_i = ASC_i + \beta_1X_{1i} + \beta_2X_{2i} + ... + \beta_nX_{ni} \qquad \text{(Equation 6.2)}$$

In this case the ASCs can measure order effects (eg if there is a preference for picking the leftmost alternative), but are otherwise unrelated to the alternatives. It is common to exclude ASCs in generic formats. Where survey participants are asked to choose between alternatives for a pre-specified type of car trip (commute to work), but not to choose between different types of trip or travel mode, generic attributes are relevant.

Attributes may have non-linear effects. For example, the marginal utility of a minute of extra travel time may be low for short trips, but high for long trips. As noted in section 4.2.7, if nonlinear effects are expected for a certain attribute, then more than two levels need to be used for this attribute to be able to estimate these nonlinearities. There are two broad approaches to accounting for such effects, either adding variables to the utility function, or using transformed variables. Dummy or effects coding of different attribute levels accommodate the former approach, adding an attribute into the utility for each level except for an excluded base level, but are practical only for a small number of attribute levels. In the transformation approach the utility function is specified as:

$$U_i = ASC_i + \beta_1f(X_{1i}) + ... \qquad \text{(Equation 6.3)}$$

Where $f(X_{ki})$ is some transformation of attribute k for alternative i, commonly of polynomial, logarithmic or exponential form, that may entail estimation of additional parameters. Modelling non-linear effects has the potential to dramatically increase the complexity of utility models and of experimental design, so should be used sparingly. Non-linearities should be justified on the basis of theory or prior empirical evidence.

Attributes related to the decision maker (age, income, children in the car, etc) can influence choices. However, unless they are specified in the design (eg with/without children), they are invariant across alternatives. Choices depend on the differences in utility between alternatives. Adding generic personal attributes to the utility functions of all alternatives would have no effect on utility differences. Therefore, personal attributes must be excluded from at least one alternative.

6.5 Efficient design

There are different approaches to experimental design which estimate the number of choice tasks (and the number of respondents required) so parameters can be estimated with the greatest expected reliability (lower expected standard errors) (Bliemer and Rose 2006). Full factorial designs include all possible combinations of attributes and levels of those attributes, but these lead to too many questions for a single respondent, or too many respondents required; in addition, the survey may ask about impossible or 'useless' choice situations, eg combinations which are very different from the usual experience of the respondent. Orthogonal partial-factorial designs reduce the number of choices, while still ensuring there are no correlations between attribute levels. However, there may still be too many choices.

We would like to estimate the unknown parameters of the utility function (and hence WTP for trip attributes) as accurately as possible. However, presenting all possible combinations of trip characteristics would lead to an overwhelming number of scenarios for participants to evaluate. An efficient design seeks to reduce the number of scenarios while still estimating the utility parameters with a high degree of accuracy.

For an efficient design, some 'useless' combinations can be eliminated by using historical trip data to narrow the choice sets. Efficient experimental design then determines a choice set which seeks to achieve significant variance between values (so people can better distinguish between the choices) but where there is minimal correlation in changes in these parameters, eg increasing the trip duration does not always increase the crash risk. Mathematically this is achieved by minimising the D-error or the A-error (Rose and Bliemer 2009; Scarpa and Rose 2008):

$$D\text{-}error = det(\Omega(\beta, x))^{1/K}$$
$$A\text{-}error = tr(\Omega)/K \qquad \text{(Equation 6.4)}$$

Where Ω is a variance-covariance matrix of the unknown parameters of the utility function being estimated (β) using experimental design x and *K* is the number of parameters (size of the variance-covariance matrix). In these equations, det refers to the determinant[30] of the matrix and tr is the trace[31].

Essentially what the D-error and A-error do is translate the expected accuracy (ie the variance) of the estimated utility function parameters into a single number that depends on the experimental design. The

[30] The determinant is the sum of product terms made up of elements from the matrix, each product term consists of n elements from the matrix, and each product term includes one element from each row and one element from each column. The number of product terms is equal to n! (where n! is n factorial) http://stattrek.com/matrix-algebra/matrix-determinant.aspx

[31] The trace is the sum of the elements on the main diagonal of the matrix

design can then be adjusted to minimise these errors, which is equivalent to designing the experiment in a way that estimates the utility parameters as accurately as possible.

The purpose of the current study was estimation of various money values, particularly the value of a statistical life. Scarpa and Rose (2008) introduce another efficiency measure relevant in this context. C-efficiency can be used to derive designs that minimise error in any particular monetary value (eg VSL), or to minimise the sample size required to achieve the desired significance level for all WTP measures. Kerr and Sharp (2010) derive the minimum sample size for C-efficiency in a model specified with a linear utility function as:

$$N_i = t^2_{\alpha/2} Var(WTP_i) WTP_i^{-2} \quad \text{(Equation 6.5)}$$

Where:

$WTP_i = -\gamma\beta^{-1}$

$Var(WTP_i) \approx \beta^{-2}(Var(\gamma) - 2\gamma\beta^{-1}(Cov(\gamma,\beta) + \gamma^2\beta^{-2}Var(\beta))$

γ is the coefficient on the attribute of interest

β is the coefficient on cost (the marginal utility of money).

Given the design it is possible to estimate Ni. Consequently, N_i can be minimised by searching over the set of possible designs. Identification of the most efficient design depends on several factors, including the purpose of the study, and prior information about the nature of the utility function, the expected coefficients and their variance, and the degree of heterogeneity in the study population. The better information is about these matters prior to development of the experimental design, the better is the expected outcome. Commonly, much of this information is unavailable when the experimental design is created. Information from previous studies and pre-testing and piloting is extremely valuable to inform development of the design.

The efficient design for this study was developed using Ngene software.[32] Ngene can optimise experimental design for C-efficiency for only one measure of WTP. Since the study sought to identify several WTP measures, which are developed from mathematical manipulation of the utility function coefficients, D-efficiency was targeted. Priors for design estimation were derived from beta-test results and various other estimates of WTP for attributes, including the current New Zealand VSL and values of injuries and travel time. To enable identification of both fatality and injury values there were two sample splits with different injury rates relative to fatality rates. Design efficiency was estimated with separate models for the two splits and with a pooled model. Results were essentially indistinguishable, so the pooled model was used to create the final experimental design. The pilot test experimental design is reported in appendix B.

6.6 Process

The pilot test comprised 72 responses, including 22 completed face-to-face at a central Auckland location and 50 completed online with respondents around the rest of New Zealand. Respondents included a mix of gender, ethnicities, ages and socio-economic backgrounds. Face-to-face respondents were recruited by phone, then interviewed by an experienced interviewer who read the background information, explanation text and questions, with the exception of the choice set questions which respondents

[32] www.choice-metrics.com

answered themselves while being observed. Respondents were offered an incentive to encourage participation and as *koha* (gift) for their time.

The lessons learned from the pilot test, and the results, are discussed in the next section. This includes additional feedback on the respondent experience, the data obtained and a discussion of how the data can be used to derive quantitative values for several attributes.

7 Results

7.1 Comprehension

Most of the commentaries recorded here are from the face-to-face interviews, as without an interviewer present, online observations rely on respondents self-reporting. The number of respondents online who had comprehension issues was potentially higher than observed.

7.1.1 Number of attributes and choice sets

The experimental design contained 60 choice tasks. To limit respondent burden, each respondent faced a subset of 10 of those tasks. The full set of tasks was allocated into six blocks as part of the experimental design process. No respondents said they felt there were too many attributes or choice sets and none were observed face-to-face as losing concentration due to the number of questions provided. However, based on the assessments of the earlier versions of the questionnaire, which had more choice sets and more attributes, the current number of 10 choice sets with seven attributes per route definition appears to be a good combination for the general population audience. Increasing complexity or length beyond this is likely to reduce the quality of engagement.

7.1.2 Differentiation in choices and attribute levels

Five respondents (three face-to-face and two online) said it was either difficult to decide between the two routes presented as they were too similar, or the attribute levels in the choice questions were too similar. It seems some respondents will round attribute levels and are reasonably insensitive to minor differences. While this may not affect the statistical modelling it can make choices harder from the respondent's point of view. Some respondents said they used the preference scale to indicate when a decision was difficult to make by indicating moderate preference instead of strong preference.

7.1.3 Most/least important attributes

Respondents were asked at the end of the survey which, if any, attributes they did not consider when making choices. Risk of injuries, risk of fatalities and cost were most commonly mentioned. About 40% of respondents (30 from 72) indicated they did not consider minor injuries, while travel time was least likely to not be considered (table 7.1). Across all three drafts, travel time has consistently been mentioned as the most important attribute for respondents.

Table 7.1 Attributes not taken into account

Attribute not considered	Count
Minor injury	30
Cost	21
Serious injury	20
Fatalities	18
Lateness	15
Heavy traffic	10
Travel time	8

Only 19 out of 72 respondents (26%) said they considered all attributes when choosing between the routes, suggesting a pattern of preference learning by the majority of respondents. This includes both respondents who ignored an attribute entirely and those who read the attribute but found it did not affect their decision. When asked, two of the 22 face-to-face respondents admitted to ignoring at least one attribute altogether. Online, respondents were not asked whether attributes were ignored; however, we know, in the absence of an interviewer, respondents online spent less time on all questions overall, suggesting similar preference learning behaviour.

Respondents who stated they did not consider at least one attribute were most likely to consider all but one attribute (20 considered all but one attribute); however, a further 10 chose to not consider two attributes and a further 15 did not consider three attributes (table 7.2).

Table 7.2 Number of attributes not considered

Number of attributes not considered	Count
1 not considered	20
2 not considered	10
3 not considered	15
4 not considered	4
5 not considered	3
6 not considered	1

Injuries, fatalities and cost were displayed towards the bottom of the attribute choice table and were more likely to be mentioned as an attribute not considered. These attributes were less important to some respondents, with most of them giving a reason as to why. However, it is possible their position on the table influenced the number who chose not to consider them. For future rounds of surveying, we suggest randomising the order the attributes are displayed for each respondent, to mitigate any potential attribute order bias.

7.1.4 Making assumptions about attributes

To counteract respondents associating crashes with delays in the earlier testing, the words ‘accidents’ or ‘crashes’ were removed and replaced with ‘injuries’ or ‘fatalities’.[33] An explanation was also added saying information about traffic or delays are to be treated separately from injury or fatality information, resulting in no pilot respondents saying they used injuries or fatalities as an indication of delays.

Although comprehension has improved overall, some respondents still add their own interpretation as to how attribute levels will affect journeys. For example, a respondent chose the slowest route for long distance trips, assuming traffic would be moving slower making it easier to look at road signs. Another chose the slower route, saying ‘I believe time adds to the sense of urgency and can then frustrate other drivers into making poor driving choices’. This same respondent said they did not look at the injury or fatality information, instead solely using travel time as an indication of safety.

7.1.5 Mentally changing values presented

As in the earlier testing, some respondents indicated they did not consider one or more attributes because they felt they could influence the numbers presented. Ten respondents commented they did not consider fatalities, injuries or lateness because their behaviour would change the numbers presented. For example, ‘as a safe driver it was less relevant to affect my decision making’ or ‘for fatalities, I trust myself to drive

[33] This also required different scaling in the calculation of the VoSL

safely, so believe other people trust themselves too'. Others chose not to consider lateness because 'I always leave a bit earlier than the time I should, to take into account of anything happening on the way' or 'I'm on flexible hours'.

Before starting the choice set questions, five of the face-to-face respondents talked unprompted about how their behaviour would change the values presented and interviewers were able to correct this assumption by explaining that, for example, crashes would be out of their control. This suggests, without an interviewer present, respondents would be more likely to make assumptions about their ability to influence the attribute levels.

Reflecting how some people process risk of harm, as well as some respondents assuming deaths and injuries would not affect them due to their behaviour, other respondents either did not believe the numbers provided, or felt the number was out of their control and therefore did not consider them. For example, 'just because deaths have happened in the past doesn't mean they'll happen again', 'what has to happen will happen' and 'people can die or get injured in their own driveways'.

In face-to-face interviews, it was explained the injury and fatality risks were not controllable. It might be necessary to present this information clearly to online respondents also. Similar comments apply to lateness also, ie the issue of concern is unexplained lateness rather than lateness that can be identified before the rote is selected, and thus avoided.

7.1.6 Thoroughness of reading

As with the beta test, respondents spent less time reading questions or introductory text in the online survey when compared with the face-to-face survey version. This was more pronounced in the pilot study and was partly due to interviews taking longer in general when done face-to-face, partly because interviewers spent time ensuring explanations were understood or repeating information when necessary. On average, online surveys took 10 minutes and face-to-face surveys took 20 minutes. Eleven online respondents completed the survey but were discarded because they completed it in under five minutes (with some as short as two or three minutes) which was deemed too fast to be able to fully read and comprehend everything. This means 18% of online responses were discarded due to 'speeding' behaviour. For a survey like this, which requires more attentive participation, face-to-face interviews are likely to provide more focused and thorough respondent engagement.

Interviews can introduce bias through interviewer cues. These are well recognised and addressed through standard practices. Interviewer use of standardised video information can help eliminate interviewer cues.

7.1.7 Crash frequency

In the pilot, the risks of injuries and fatalities were provided per 100 billion kilometres travelled, whereas the alpha and beta tests provided a number per year. Face-to-face in the pilot, respondents were given a show card explaining the crashes and just one of the 22 respondents noticed and asked questions about the phrase 'per 100 billion kms travelled'.

Once interviewers explained what is meant by 'per 100 billion kms travelled' approximately half the face-to-face respondents either asked for more explanation, or the interviewer felt their response meant it was required. Two face-to-face respondents talked about the number of 'deaths per year' later in the survey and another said he was thinking about a total number of deaths or injuries on the road and did not tie that back to either a time period, or distance.

This suggests, without the assistance of an interviewer, it is likely that respondents will not pay full attention to the frequency or fully understand what is meant. Poor comprehension of this phrase suggests using the phrase ‘per 100 billion kms travelled’ is not only affecting the respondent’s ability to answer the question but is also not being correctly interpreted by some.

7.1.8 Trip purpose

Prior to the pilot study, respondents were provided with a trip purpose (such as work, holidays or to visit friends and family). This was added to the survey during the alpha test due to feedback from respondents but was subsequently removed after the beta test. Trip purpose is known to have an effect, but is common across the two choices in any one choice task. Hence it can be addressed by design replicates for different trip purposes. Sample size limitations led to the trip purpose remaining excluded from the pilot study, although the intention is for purpose to be part of the design of the final study. A larger sample size will allow adoption of an experimental design to account for differences in values between trip purposes. Exclusion of trip purpose from the pilot resulted in anticipated responses that said choice would change depending on the trip purpose. For example, one respondent said they answered all questions as if they were travelling for work, and therefore chose the fastest route. They indicated their response would be different if they were responding for a less urgent trip. We do not know what type of trip the majority of respondents are imagining in the pilot study. We recommend inclusion of trip purpose in the final study through a split sample design.

7.1.9 Realism of the choice set tasks

In all three rounds of testing, respondents commented they were able to answer the questions provided; however, some said their choices were not what they would do in real life. Hypothetical bias was experienced in the pilot, for example, ‘in the real world you’d use the faster route on google’ or ‘You don’t usually have accident statistics when choosing what route to take’. In the survey, we currently acknowledge ‘this is more information than you usually have but we would like you to consider it as if it was real’; however, at least three of the respondents face-to-face said they were unable to consider all the information as if it was real, because either they would not have this information in real life or they would not make that type of decision in real life.

7.1.10 Effects of survey mode

7.1.10.1 Comprehension issues

The majority of comprehension issues described were observed during the face-to-face survey, when interviewers were available to answer queries, were actively watching for signs a respondent had misunderstood and could ask a respondent questions about their responses. The ability to look for comprehension issues meant these issues were often remedied by the interviewer. For example, if respondents said they were a safe driver and would not be involved in a crash, the interviewer explained how fatalities or injuries were also caused by the condition of the road or the behaviour of other drivers. Instead, those online who commented they felt they were a safe driver and would not be involved in a crash, only did so after completing the choice set questions. This was the case with all other comprehension issues observed throughout the testing.

7.1.10.2 Thoroughness of reading

The significant proportion of online responses discarded due to ‘speeding’ behaviour and the difference in interview length between the modes indicated online respondents spent less time reading questions thoroughly. Some face-to-face respondents indicated they wanted to complete the survey faster, but interviewers were able to manage the pace and did not skip over screens or rush through explanations. Interviewers probed for more information or clarity when it was clear a respondent had either misunderstood or had not fully listened to what was required by a question. In the beta test, face-to-face respondents who completed an accompanied interview were able to more clearly articulate how they had responded and why they had chosen one route over another. This was more pronounced again in the face-to-face interviews from the pilot when compared with the face-to-face accompanied interviews of the beta test.

SP studies require a higher level of attention and consideration when compared with most other surveys of the general public, as well as a degree of learning or coming up to speed with the background information about how we want people to respond to each question. Respondents in the face-to-face survey in the pilot were observed to be more attentive and thorough in their consideration of each choice set when compared with the online respondents or the face-to-face accompanied respondents in the alpha or beta tests, suggesting face-to-face interviews will provide higher quality data, with less opportunity for respondents to misunderstand or skip questions and explanations.

7.2 Observations

7.2.1 Introduction text

Online respondents spent less time on the first introduction screen than face-to-face respondents, meaning it was likely some online respondents did not read the full introduction, and two face-to-face respondents commented unprompted that the introduction was long. Our observation is, if respondents feel the introduction is too dry, wordy or unnecessary to fully understand the purpose of the research, it sets the tone for the rest of the survey, causing respondents to assume other screens have superfluous information and speed through those screens, or skip them entirely.

In the beta test, some text was moved to a second screen to view if respondents required more information. Seven pilot respondents clicked this for more information, suggesting those who want to know more will click this link. We suggest moving more information to this second ‘if needed’ screen and simplifying the introduction to something like:

> *This research is for the New Zealand Transport Agency and will help them to better understand how people decide which route they will drive when given more than one option. This information can be used to help make decisions about policy or infrastructure.*

7.2.2 Recent trips and recent trip length questions

When asked about their typical long-distance or irregular local trips, some face-to-face respondents were observed to be confused about whether they were meant to include work trips or how long a trip must be to count as a long-distance trip. Three respondents provided a longer trip length for their local trip than their long-distance trip.

We suggest removing these questions to avoid creating confusion in the survey.

7.2.3 Crash experience

Respondents were asked if they have personally been involved in or know anyone else who has had any experience with crashes resulting in different injury types. Approximately half the face-to-face respondents asked questions about either the timeframe or how close their relationship must be to the person involved in a crash. We suggest providing a timeframe such as ‘within the last five years’.

7.3 Numerical results and analysis

7.3.1 Data

Data from the survey was verified and converted to numerical format to enable mathematical analysis. This was a pilot study in which the key issue was whether the survey results could be used to estimate reliable relative values for the selected transport impacts: value of travel time, lateness, congestion and risks of injuries and fatalities. The magnitudes of derived values are not relevant in this process. The data analysis in this section examines this more limited functional question. Survey responses were converted to a format suitable for numerical analysis using NLOGIT6, a software package developed for CM.[34]

The data contained responses from 83 individuals who had each completed 10 choice tasks (or scenarios), as prescribed by the statistical design (appendix B).[35] Participants were allocated to one of six design blocks, with 13 to 15 responses received for each block. As noted in section 7.1.6, 11 individuals completed the choice tasks in under five minutes, and it was assumed those individuals did not give sufficient consideration to the choice tasks. Consequently, analysis proceeded with two different groups of respondents:

1 The *full sample* included all 83 respondents

2 The *reduced sample* of 72 respondents excluding the 11 individuals who completed the survey in less than five minutes.

The survey asked respondents for their strength of preference in making a choice (figure 6.1). This was to help respondents make choices. However, for analysis we treated this as a simple binary choice, with preference expressed for either route one or two.

7.3.2 Route purpose

In the alpha and beta tests, route purpose was provided to help respondents to better imagine the journey. Respondents stated they considered the purpose in making their route preferences and in weighing up the differences. In the pilot study, and reflecting the small sample size, no route purpose was provided. However, because preferences were purpose dependent, we suggest it is introduced again in a full survey.

For generic choices an alternative–specific constant is not normally required when modelling differences in utility between alternatives. However, to accommodate the possibility of order–effects (eg some respondents may be more likely to select the left-most choice, or vice versa), models with and without alternative–specific constants (dummy variables) were estimated. Hence, for each model type, four models were estimated:

- *full sample* with and without an alternative–specific constant
- *reduced sample* with and without an alternative–specific constant.

[34] www.limdep.com/products/nlogit/

[35] For the online survey, 6,200 invitations were sent, 477 responded to the invitation, of which 154 were eligible, 82 only partially completed the survey and 61 completed it. For the face-to-face interviews 32 contact attempts were made, three did not respond, three were subsequently not available on the day, and two were ‘no-shows’; 22 completed the survey.

7.3.3 Model type

Once the data is collected, different techniques can be used for analysis, as described in section 3.4. Several different model types were fitted:

- multinomial logit model (MNL), as described in section 3.4.1
- scaled multinomial logit model (sMNL)
- latent class multinomial logit model (LCMNL)
- RPL model, also known as the mixed logit (ML) model, as described in section 3.4.2.

This is not a complete range of potential models, but allows some sensitivity testing with respect to model type, and provides confidence the data is suitable for a range of model types. All models converged and permitted estimation of willingness to pay for attribute changes. Further model testing with the collected data would be advantageous in the final study:

- MNL models provide checks on data integrity before proceeding to more complex models.
- sMNL models allow for some respondent heterogeneity, in terms of variance of the individuals' responses, but assume a common underlying preference structure. Whereas MNL treats each response as independent and each individual as identical, sMNL uses the panel data to recognise differences between individuals, and assesses the consistency of responses by each individual. Some individuals' responses to the 10 choice scenarios were very consistent, whereas others' responses were inconsistent. sMNL puts more weight on responses from the more consistent individuals.
- LCMNL models incorporate limited respondent heterogeneity. They use the panel data to characterise groups of respondents with similar response patterns. Within these groups each individual is treated as having identical preferences, but preferences differ between the groups. LCMNL models identified significant respondent heterogeneity. Because LCMNL identifies groups of individuals with similar preferences, whereas RPL allows each individual to have unique preferences, LCMNL models have less ability than RPL models to accommodate respondent heterogeneity.
- RPL models were the focus of final analysis. The RPL models were estimated with normal distributions of the parameters, another matter for sensitivity testing with a larger sample. The four RPL models are reported in table 7.3.

All parameter values are statistically significant at better than the 1% level, apart from the constant in model 1, which is significant at the 1.5% level.

The *Rho*^{2} measures indicate reasonably good model fit, equivalent to R^2 of about 0.5 in a linear regression model (Hensher et al 2005). Higher *Rho*^{2} scores are preferred, favouring model 2, although the differences are very small. Smaller normalised information criteria scores are preferred, again indicating model 2 as the preferred alternative. Models 1 and 2 include significant alternative-specific constants, indicating the presence of an order effect. Despite smaller sample size, the *reduced sample* models (models 2 and 4) offer better model fit, supporting exclusion of the rapid completers. However, it is extremely encouraging the four models have very similar coefficients with high levels of significance, and very similar model fit statistics, together indicating a lack of sensitivity to assumptions about exclusions from the final sample and inclusion of an alternative–specific constant.

Table 7.3 Random parameters logit models (p values in parentheses)

	Model 1 Full sample	**Model 2 Reduced sample**	**Model 3 Full sample**	**Model 4 Reduced sample**
Constant	.236 (.0150)	.296 (.0066)	-	-
Travel time	-.056 (.0000)	-.057 (.0000)	-.054 (.0000)	-.054 (.000)
Lateness	-.033 (.0018)	-.042 (.0008)	-.030 (.0033)	-.039 (.0009)
Heavy	-.029 (.0000)	-.031 (.0000)	-.028 (.0000)	-.030 (.0000)
Cost	-.165 (.0000)	-.165 (.0006)	-.157 (.0001)	-.158 (.0004)
Death1	-.653 (.0000)	-.640 (.0001)	-.628 (.0000)	-.682 (.0000)
Death2	-.454 (.0000)	-.469 (.0001)	-.439 (.0000)	-.458 (.0000)
Variance travel time	.044 (.0000)	.047 (.0000)	.043 (.0000)	.044 (.0000)
Variance lateness	Fixed	Fixed	Fixed	Fixed
Variance heavy	.020 (.0005)	.022 (.0008)	.020 (.0005)	.024 (.0002)
Variance cost	Fixed	Fixed	Fixed	Fixed
Variance death1	.536 (.0001)	.574 (.0452)	.507 (.0001)	.545 (.0002)
Variance death2	.372 (.0001)	.409 (.0004)	.361 (.0001)	.397 (.0002)
$N = N_{individuals}$	83	72	83	72
Number of choices	830	720	830	720
$K = N_{parameters}$	11	11	10	10
Model fit measures				
Log-likelihood	-448.687	-386.838	-451.711	-389.005
Rho^2	.208	.211	.204	.207
AIC/N	1.108	1.105	1.113	1.108
AIC3/N	1.121	1.120	1.125	1.122
CAIC/N	1.121	1.120	1.125	1.122
BIC/N	1.129	1.118	1.124	1.120
aBIC/N	1.102	1.097	1.107	1.101

Rho^2 = Adjusted McFadden's R^2; AIC = Akaike information criterion; AIC3 = Bozdogan AIC; CAIC = consistent AIC; BIC = Bayesian Information Criterion; aBIC = adjusted BIC

7.3.4 Split sample design and derivation of VoSL

Risks of fatality and of two levels of injury were included as attributes in the choice experiment alternatives (figure 6.1). To reduce participant burden, and to reduce the statistical design to a manageable number of choice events (particularly for the small sample pilot), it was decided to not include varying attribute levels for each of the two injury types in the pilot survey design. It would have been unrealistic to include risks of fatality, but not of injury. To enable identification of VoSL while including injuries a split-sample approach was used. In each of the two sample splits the rate of serious and minor injuries was fixed in proportion to the fatality rate, but this proportion varied between the splits:[36]

- In split one, there were five serious injuries and 35 minor injuries for every fatality.
- In split two, there were nine serious injuries and 63 minor injuries for every fatality.

The estimated models contained two variables relating to fatality and injury risk, ie DEATH1 for split one (with associated coefficient α_1) and DEATH2 for split two (with associated coefficient α_2). The attribute DEATH1 was set to zero for individuals in split two, and DEATH2 was set to zero for individuals in split one. Since route choice depends on utility *differences* between the two routes, DEATHi does not affect choices for split $j \neq i$. Because injuries were not part of the statistical design, but their rate relative to fatalities was, the value of fatalities and injuries is confounded in the results, requiring some mathematical manipulation to extract the value of a fatality. The following section describes the method used.

Let:

X_i = serious injuries per fatality for split i

W_i = minor injuries per fatality for split i

$W_i/X_i = k = 7$ in this study

$X_2/X_1 = W_2/W_1 = y = 1.8$ in this study

R_i = value of combined injuries and fatality in split i (per fatality)

V_f = value of a fatality

V_S = value of a serious injury

V_M = value of a minor injury

Then:

$$R_i = V_f + X_iV_S + W_iV_M = V_f + X_i(V_S + kV_M) = V_f + X_i(q) \quad \text{(Equation 7.1)}$$

$$R_2 - R_1 = (X_2 - X_1)q$$

Rearranging:

$$V_f = R_1 - ((R_2 - R_1)/(y - 1)) = (yR_1 - R_2)/(y-1) \quad \text{(Equation 7.2)}$$

In a linear utility function, R_i is the ratio of estimated model coefficients: $R_i = -\alpha_i/\beta_i$

These coefficients are estimated in the same model via dummy coding of the splits, so $\beta_1 = \beta_2 = \beta$

$$V_f = (\alpha_2 - y\alpha_1)/(\beta(y-1)) \quad \text{(Equation 7.3)}$$

In the pilot study, y = 1.8, so

$$V_f = (1.8R_1 - R_2)/(0.8) = (\alpha_2 - 1.8\alpha_1)/(0.8\beta) \quad \text{(Equation 7.4)}$$

[36] The values are the (broadly rounded) observed ratios for serious/minor injury in New Zealand using data provided by Wayne Jones (Ministry of Transport – pers comm) and in 'Road safety outcomes'. https://nzta.govt.nz/assets/resources/road-safety-outcomes/docs/regional-quarterly-outcomes.xls. There were six or seven serious injuries per fatality in 2014–2016. There were nine hospitalisations per fatality over the same period. We needed to vary the serious/fatal ratio to identify the model, resulting in five and nine used in the pilot.

The value of a fatality is a function of three model coefficients (α_1, α_2, β). Consequently, the variance of V_f is a function of the variance of those same three coefficients.

Previous New Zealand studies of VoSL (see section 2.1.4) have not reported confidence intervals. However, it is critical to know how reliable WTP estimates are, both in policy analysis applications, and in identification of the required sample size for a given level of accuracy in the final study.

Because, as shown above, WTP for each attribute is derived algebraically from the statistical model outputs (which are themselves uncertain), variance of WTP is not a direct output from the statistical model, but must be estimated through simulation. The simulation process entailed creation of 10,000 different estimates for each parameter in the WTP equation (α_1, α_2, β). These parameter estimates were drawn randomly from the distributions for α_1, α_2 and β, taking account of their covariances. V_f was then estimated for each (α_1, α_2, β) combination to generate a distribution for mean V_f.

This approach does not permit direct derivation of WTP to avoid fatality risk. To derive the value of a statistical life, rather than the value of a fatality rate, V_f was scaled to account for exposure to risk. This was done by multiplying by average journey distance in the choice experiment (160 km)[37] and dividing by 100 to derive dollars per life. WTP estimates are reported in table 7.4.

For all attributes except VoSL, mean WTP estimates were simply the ratio of the coefficient associated with that attribute (α_i) and the cost coefficient (β).

Because WTP measures rely on the estimated distribution of the statistical model parameters themselves, and the spreads of those distributions have inverse relationships with sample size, parameter estimates are relatively widely dispersed for the small sample in the pilot study, resulting in broad confidence intervals for WTP. Bigger samples in the final study will provide narrower confidence intervals (see discussion in section 7.3.5). Confidence interval calculations were undertaken in the NLOGIT6 package.

Table 7.4 Willingness to pay estimates (standard errors in parentheses)

	Model 1 Full sample	Model 2 Reduced sample	Model 3 Full sample	Model 4 Reduced sample
Travel time ($/hour)	20.34 (3.24)	20.70 (3.66)	20.70 (3.42)	20.64 (3.78)
Lateness ($/hour)	12.00 (2.40)	15.36 (2.76)	11.52 (2.52)	14.88 (2.82)
Heavy traffic ($/percent)	0.174 (.037)	0.189 (.044)	0.176 (.039)	0.190 (.047)
Statistical life ($m/life) 95 % ci	8.75 (3.01) 2.86 – 14.64	8.29 (4.14) 0.16 – 16.42	8.82 (3.07) 2.80 – 14.85	9.76 (3.65) 2.61 – 16.91

As with the model coefficients, monetary value estimates are reasonably uniform across the four models, with no statistically significant differences between any of the estimates. Standard errors are small relative to the mean for travel time, lateness and heavy traffic. The mean value of travel time (using model 2 as an example) is $21 per hour, with a 95% confidence interval of $17 – $24 per hour. This confidence interval is expected to contrast significantly with a meaningful expansion in sample size from the 72 individuals in model 2.

[37] Because of the small sample size, average values were used, rather than splitting the data into long and short journeys. This was done to provide proof of concept, ie that the methodology can be used to derive values, and has been used to produce illustrative values only. In a larger survey the data could be analysed separately to understand the influence of journey length and/or trip purpose.

The mean value of a statistical life (model 2) is $8.3 million, with a 95% confidence interval of $0.2 – $16.4 million. Confidence intervals for the value of a statistical life are broad for two main reasons. First, the sample size in the pilot is small. Second, the split sample approach means variance from three associated coefficient estimates contribute to the dispersion of the distribution of the mean, rather than the two coefficient variances in the other WTP estimates. Note there are many caveats on this estimate, particularly around journey purpose and trip length, so it should not be used for policy purposes. It does, however, illustrate the viability of the process.

7.3.5 Sample size

To examine the sample size which might be used for a full survey, we analysed the impacts of the sample size on error levels (table 7.5). For the current sample (N=72) the standard error of the mean (SEM) estimated VoSL is $4.15 million, yielding a 95% confidence interval of +/- $8.13 million (98% of the mean value). Increasing the sample to 2000 respondents would narrow the range down to +/- $1.54 million (19% of the mean value). Note these estimates do not account for the impact of differences in values associated with different trip purposes, which were not tested in the pilot study. The mean values are derived from the analysis and we have no basis for estimating a different mean for a different sample size, although in practice, the mean might change with a change in sample size. However, because standard errors for each of the estimated coefficients in the equation that derives VoSL and other values are inversely proportional to the square root of the sample size it is possible to estimate the resulting confidence interval from any sample size.[38]

Table 7.5 Implications of sample size for error levels

N:	72	500	1,000	2,000	5,000	10,000	20,000
Mean VoSL from Monte Carlo Analysis (MCA) ($m)	$8.29	$8.29	$8.29	$8.29	$8.29	$8.29	$8.29
SEM adjusted for sample size ($m)	$4.15	$1.57	$1.11	$0.79	$0.50	$0.35	$0.25
95% error ($m)	$8.13	$3.08	$2.18	$1.54	$0.98	$0.69	$0.49
Error %	98%	37%	26%	19%	12%	8%	6%
Mean value of travel time from MCA ($/hour)	$20.70	$20.70	$20.70	$20.70	$20.70	$20.70	$20.70
SEM adjusted for sample size ($/hour)	$3.66	$1.39	$0.98	$0.69	$0.44	$0.31	$0.22
95% error ($/hour)	$7.17	$2.72	$1.92	$1.36	$0.86	$0.61	$0.43
Error %	35%	13%	9%	7%	4%	3%	2%
Mean value of lateness from MCA ($/hour)	$15.36	$15.36	$15.36	$15.36	$15.36	$15.36	$15.36
SEM adjusted for sample size ($/hour)	$2.76	$1.05	$0.74	$0.52	$0.33	$0.23	$0.17
95% error ($/hour)	$5.41	$2.05	$1.45	$1.03	$0.65	$0.46	$0.32
Error %	35%	13%	9%	7%	4%	3%	2%

[38] In the table, the values are calculated as follows, using a sample size of 1,000 and VoSL as an example: SEM_{1000} = $4.15 × SQRT(72) ÷ SQRT(1000). 95% error = $1.11 × 1.96 = $2.18. And error = $2.18 ÷ $8.29 = 26%

8 Conclusions and suggestions for a full survey

8.1 Objectives

The objective of this study focused on testing practicality and feasibility of methods. It was not to derive an estimate of VoSL, the value of travel time, or other values of road use. Rather it was to test whether it would be possible to estimate reliable relative values of these impacts using a single CM survey which included several different attributes.

Following two rounds of survey testing, and then using a small survey and analysis of the data collected, this study has confirmed it is possible to use a CM survey to estimate VoSL, the value of injuries and the value of time. This suggests it is worthwhile proceeding with the design and implementation of a larger survey with more participants.

Although the small survey described in this study has been used to produce estimates of values (section 7.3), the small survey size means the error margins are large. The values should not be used to draw conclusions about the mean values which might be derived from a future, larger survey.

8.2 Success of experimental design

All pilot study participants were willing and able to answer all 10 choice questions presented to them. The experimental design was successful: all model coefficients were estimable with reasonably high precision,[39] even with a very small sample. The importance of preference heterogeneity was underlined by significant improvements in model fit in models that accommodated heterogeneity, most notably the RPL model. Derivation of confidence intervals for the value of a statistical life is an advance over previous New Zealand studies, which have simply reported estimated mean values. The broad confidence intervals for estimated mean values of a statistical life were anticipated in this small pilot study and are not a matter for concern. Knowledge of expected mean monetary values and their distributions will be valuable for improving experimental designs in subsequent studies, which together with a much larger sample size in the final study will deliver significant gains in precision of benefit estimates. A larger sample size would also allow analysis of the influence of other factors on the results, eg the influence of age, income or trip purpose.

8.3 Survey mode, design and wording

A number of suggestions are made regarding the survey mode and its wording.

- Survey mode: It is important to the choice game process that the context of the choice and the nature of each attribute is fully understood. Face-to-face interviewing ensures more reliable and controlled responses compared with a self-completion online survey. However, it is considerably more expensive. The statistical analysis suggested respondents produced reasonable results in the online format, although the samples were too small to identify significant differences between modes. Other hybrid modes could be considered (see below).

[39] They are precise in the sense they have low p-values (table 7.3)

- Design: Experiments with approximately seven attributes appear to provide a good balance between maximising the values derived from a survey and complexity for survey participants.
- Risk information: The pilot survey and the previous tests experimented with different ways to provide risk information, balancing the need for this to be realistic with that of making it useful for respondents making choices. The pilot provided information on injuries and fatalities per 100 billion kms travelled. This was used as a proxy for the uncontrollable risk, was difficult for respondents to relate to, but was helped by face-to-face explanations. This appears to be a sensible approach.
- Other mixed-mode options: Other hybrid or mixed-mode options could be considered, such as a hybrid phone and online survey or an online survey plus video. The first (less preferred) could involve a phone interview to explain the concepts to a respondent and to ask all non-choice set questions before asking them to complete an online self-completion choice set questionnaire. The second (preferred) option is an online survey in which a respondent first watches a recorded video of an interviewer explaining the survey and concepts such as how different roads affect crash numbers before they can complete the choice set questions. These options could go some way towards emulating a fully intensive face-to-face experience.
- Practice choice set: For face-to-face interviewing, we recommend adding a trial choice set, after which the interviewer asks questions about the response process to assess comprehension. Any issues can then be rectified before allowing the respondent to continue with the remaining choice questions.
- Explanation about fixed attributes: To avoid assumptions that either an attribute does not apply to a respondent, or they can influence the values presented by their behaviour, more explanation could be included to reiterate that values are fixed for both routes and cannot be influenced.
- Trip purpose: In all three drafts of the questionnaire, respondents have commented that the purpose of the trip affects how they respond, particularly with respect to the value of time. It was excluded from the pilot because of the limited sample size. However, re-introducing trip purpose to the choice sets is recommended for the final survey.
- Simplifying introduction text: To ensure people fully consider the explanations, we suggest simplifying the introduction text.

8.4 Analysis

Several models were used to analyse the data collected: MNL, sMNL, LCMNL and RPL models. All produced results which permitted estimation of WTP for attribute changes. RPL models were chosen because they were more flexible, given the differences amongst respondents. In a full survey, a similar set of models should be considered for analysis, with the decision on the best approach taken on the basis of statistical measures of model fit (such as the various information criteria scores) recognising more flexible models entail estimation of a large number of parameters and more parsimonious models can perform better. Whichever model is applied the precision of estimated mean WTP values, and hence their reliability in applied uses, must be conveyed through reporting estimated confidence intervals, which will require some form of simulation or approximation based on estimated parameter variances and covariances.

In a pilot study the analysis did not make full use of the information collected, including the age and income level of the respondent, or the previous experience of the respondent with injuries or fatalities (of friends or family). In a larger study, the influence of these factors on WTP could be examined.

9 References

Abelson, P (2008) Establishing a monetary value for lives saved: issues and controversies. *Office of Best Practice Regulation, Department of Finance and deregulation working papers in cost-benefit analysis. WP 2008-02.*

Accent and RAND Europe (2010) *Review of stated preference and willingness to pay methods*. Accessed 1 May 2018. http://webarchive.nationalarchives.gov.uk/+/http:/www.competition-commission.org.uk/our_role/analysis/summary_and_report_combined.pdf

Access Economics (2008) *The health of nations: the value of statistical life*. Canberra: Australian Safety and Compensation Council, Australian Government. 171pp.

Adamowicz, W, P Boxall, M Williams and J Louviere (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics 80,* no.1: 64–75.

Adamowicz, W, D Dupont and A Krupnick (2005) Willingness to pay to reduce community health risks from municipal drinking water, a stated preference study. In *Proceedings of Session VI: Valuing Morbidity and Mortality: Drinking Water. UN workshop Morbidity and Mortality: How Do We Value the Risk of Illness and Death?*

AEA Technology Environment (2005) Service contract for carrying out cost-benefit analysis of air quality related issues, in particular in the Clean Air for Europe (CAFE) Programme. Methodology for the cost-benefit analysis for CAFE: volume 2: health impact assessment. *AEAT/ED51014/Methodology volume 2: Issue 2.*

Agarwal, MK and PE Green (1991) Adaptive conjoint analysis versus self-explicated models: some empirical results. *International Journal of Research in Marketing 8*, no.2: 141–146.

Aldy, JE and SJ Smyth (2014) Heterogeneity in the value of life. *NBER working paper 20206*. National Bureau of Economic Research.

Algers, S, P Bergström, M Dahlberg and J Lindqvist Dillén (1998) Mixed logit estimation of the value of travel time. *Working paper series 1998:15*, Uppsala University, Department of Economics.

Alivand, M and HH Hochmair (2015) Choice set generation for modeling scenic route choice behavior using geographic information systems. *Transportation Research Record: Journal of the Transportation Research Board 2495*: 101–111.

Alivand, M, HH Hochmair and S Srinivasan (2015) Analyzing how travelers choose scenic routes using route choice models. *Computers, Environment and Urban Systems 50*: 41–52.

Alpizar, F, F Carlsson and P Martinsson (2001) Using choice experiments for non-market valuation. *Working Papers in Economics 52:* 1–37.

Anderson, WP (2011) Assessing risk and resilience for transportation infrastructure in Canada. In *Proceedings of Canadian Transportation Research Forum 46th Annual Conference*, Gatineau, Quebec, 29 May–1June 2011.

Ansolabehere, S and BF Schaffner (2014) Does survey mode still matter? Findings from a 2010 multi-mode comparison. *Political Analysis 22,* no.3: 285–303.

Antoniou, C (2014) A stated-preference study of the willingness-to-pay to reduce traffic risk in urban vs rural roads. *European Transport Research Review 6*: 31–42.

Arentze, T, A Borgers, H Timmermans and R Del Mistro (2003) Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part E 39*: 229–244.

Arrow, KJ, R Solow, E Leamer, P Portney, R Radner and H Schuman (1993) Report of the NOAA Panel on Contingent Valuation. *Federal Register 58*: 4601–4614.

Arup, ITS and Accent (2015) *Provision of market research for value of travel time savings and reliability: non-technical summary report*. UK: Department for Transport.

Australian Department of Prime Minister & Cabinet (2014) *Best practice regulation guidance note. Value of statistical life*. Canberra.

Austroads (2012) *Guide to project evaluation: part 4: project evaluation data*. AGPE04-12.

AVV (2004) *Nota mobility*. AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management.

Axhausen, K, I Ehreke, A Glemser, S Hess, C Jödden, K Nagel, A Sauer and C Weis C (2014) *FE-Projekt-Nr. 96.996/2011 Ermittlung von Bewertungsansätzen für Reisezeiten und Zuverlässigkeit auf der Basis eines Modells für modale Verlagerungen im nicht-gewerblichen und gewerblichen Personenverkehr für die Bundesverkehrswegeplanung Entwurf Schlussbericht*. IVT, ETH.

Batley, R and N Ibáñez (2009) Randomness in preferences, outcomes and tastes, an application to journey time risk, *International Choice Modelling Conference*, Yorkshire, UK.

Bateman, IJ, BH Day, AP Jones and S Jude (2009) Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management 58*: 106–118.

Baumol, WJ and WE Oates (1988) *The theory of environmental policy*. 2nd ed. Cambridge University Press.

Beck, M, T Kjaer and J Lauridsen (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment, *Health Economics 20*, no.3: 273–86.

Bhat, CR and R Sardesai (2006) The impact of stop-making and travel time reliability on commute mode choice, *Transportation Research Part B 40*, no.9: 709–730.

Bishop, RC (2003) Where to from here. Pp537–566 in *A primer on nonmarket valuation*. PA Champ, KJ Boyle and TC Brown (Eds). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bliemer, MCJ and JM Rose (2006) *Designing stated choice experiments: state-of-the-art*. Institute of Transport Studies, University of Sydney.

Bliemer, MCJ and JM Rose (2009) Efficiency and sample size requirements for stated choice experiments. *Transport Research Board Annual Meeting,* Washington DC.

Bliemer, MCJ and JM Rose (2010) Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Special issue in *Transportation Research Part B 46*, no.3: 720–734.

Bliemer, MCJ and JM Rose (2011) Experimental design influences on stated choice outputs: an empirical study in air travel choice, *Transportation Research Part A 45*: 63–79.

Bliemer, MCJ, JM Rose and S Hess (2008) Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling 1*, no.1: 98–126.

Bliemer, MCJ, JM Rose and DA Hensher (2009) Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological 43*, no.1: 19–35.

Braga, J and C Starmer (2005) Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics 32*, no.1: 55–89.

Brazell, JD and JJ Louviere (1998) *Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden*. Sydney: The University of Sydney Department of Marketing.

Brouwer, R, T Dekker, J Rolfe and J Windle (2010) Choice certainty and consistency in repeated choice experiments. *Environmental Resource Economics 46:* 93–109.

Bureau of Infrastructure, Transport and Regional Economics (BITRE) (2009) Cost of road crashes in Australia 2006*. BITRE report 118*. Canberra.

Bureau of Labor Statistics (2016) *Usual weekly earnings of wage and salary workers third quarter 2016*. News release Thursday, October 20, 2016. USDL-16-2025.

Bureau of Transport Economics [BTE] (2000) Road crash costs in Australia. *BTE report 102*.

Burke, RR, BA Harlam, BE Kahn and LM Lodish (1992) Comparing dynamic consumer choice in real and computer-simulated environments. *Journal of Consumer Research 19*, no.1: 71–82.

Cameron, TA, GL Poe, RG Ethier and WD Schulze (2002) Alternative non-market value elicitation methods: Are the underlying preferences the same? *Journal of Environmental Economics and Management 44*, no.3: 391–425.

Carlsson, F (2009) Design of stated preference surveys: is there more to learn from behavioural economics? *Working Papers in Economics 418:* 1–20.

Carlsson, F and P Martinsson (2001) Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *Journal of Environmental Economics and Management 41*, no.2: 179–192.

Carlsson, F and P Martinsson (2006) *How much is too much? An investigation of the effect of the number of choice sets, starting point and the choice of bid vectors in choice experiments.* Accessed 2 May 2018. http://economics.handels.gu.se/digitalAssets/1456/1456376_ere-2008.pdf

Carlsson, F, MR Morkbak and SB Olsen (2012) The first time is the hardest: a test of ordering effects in choice experiments. Journal of Choice Modelling 5, no.2: 19–37.

Carrion, C and D Levinson (2012) Value of travel time reliability: A review of current evidence. *Transportation Research Part A*, *46*: 720–741.

Carson, R, JJ Louviere, D Anderson, P Arabie, D Bunch, DA Hensher, R Johnson, W Kuhfeld, D Steinberg, J Swait, H Timmermans and J Wiley (1994) Experimental analysis of choice. *Marketing Letters 5*, no.4: 351–367.

Caussade, S, J de Dios Ortúzar, L Rizzi and DA Hensher (2005) Assessing the Influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B 39*, no.7: 621–640.

Cenek, P and N Jamieson (2012) Quantification of safety benefits resulting from road smoothing. *Opus International Consultants Limited Central Laboratories report 12-529B39.*

Cherchi, E and DA Hensher (2015) Workshop synthesis: Stated preference surveys and experimental sdesign, an audit of the journey so far and future research perspectives. *Transportation Research Procedia 11*: 154–164.

Chilton, S, J Covey, M Jones-Lee, G Loomes and H Metcalf (2004) *Valuation of health benefits associated with reductions in air pollution*. Department for Food, Environment and Rural Affairs.

Clough, P, J Guria and M Bealing (2015) Approaches to valuing injury and mortality risk in transport assessments. *NZ Transport Agency research report 571*. 78pp.

Colman, A (2006) Thomas C. Schelling's psychological decision theory: introduction to a special issue. *Journal of Economic Psychology 27:* 603–608

Committee on the Medical Effects of Air Pollution (COMEAP) (2010) The mortality effects of long-term exposure to particulate air pollution in the United Kingdom. *A report by the Committee on the Medical Effects of Air Pollutant*s.

Daniels, RF and DA Hensher (2000) Valuation of environmental impacts of transport projects: The challenge of self-interest proximity. *Journal of Transport Economics and Policy 34,* no.2: 189–214.

Dawes, RM (1964) Social selection based on multidimensional criteria. *The Journal of Abnormal and Social Psychology 68*, no.1: 104–109.

Day, B, IJ Bateman, RT Carson, D Dupont, JJ Louviere, S Morimoto, R Scarpa and P Wang (2012) Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management 63,* no.1: 73–91.

De Dios Ortuzar, J, D Roncagliolo and U Velarde (1997) *Interactions and independence in stated preference modelling*. Department of Transport Engineering. Pontificia Catholic University of Chile.

De Dios Ortuzar, J and LG Willumsen (2011) *Modelling transport*. 4th ed. Wiley.

DeMaio, TJ (1984) Social desirability and survey measurement: a review. Pp257–281 in *Surveying subjective phenomena: volume 2.* CF Turner and E Martin (Eds). New York, NY: Russel Sage.

Denne, T and S Atkins (2015) *Policies to reduce harmful emissions from vehicles: costs and benefits*. Report to NZ Ministry of Transport.

Department for Transport (2007) Delivering a sustainable railway. Presented to Parliament by the Secretary of State for Transport, by Command of Her Majesty July 2007. *White paper CM 7176.*

Department for Transport (2015) *Provision of market research for value of travel time savings and reliability*. Non-technical summary report. Prepared in partnership with Arup, ITS Leeds and Accent. Department for Transport, London.

Department for Transport (2016) *Reported road casualties Great Britain: 2015 annual report. Moving Britain ahead*.

Desaigues, B, D Ami, A Bartczak, M Braun-Kohlova, S Chilton, M Czajkowski, V Farreras, A Hunt, M Hutchinson, C Jeanrenaud, P Kaderjak, V Maca, O Markiewicz, A Markowska, H Metcalf, S Navrud, JS Nielsen, R Ortiz, S Pellegrini, A Rabl, R Riera, M Scasny, ME Stoeckel, R Szanto and J Urban (2011) Economic valuation of air pollution mortality: A 9-country contingent valuation survey of value of a life year (VoLY). *Ecological Indicators 11*: 902–910.

DeShazo, JR and G Fermo (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency, *Journal of Environmental Economics and Management 44*: 123–143.

Dolan, P, R Metcalfe, V Munro and MC Christensen (2008) Valuing lives and life years: anomalies, implications, and an alternative. *Health Economics, Policy and Law 3:* 277–300.

Duffy, B and K Smith, G Terhanian and J Bremer (2005) Comparing data from online and face-to-face surveys. *International Journal of Market Research 47*, no.6: 615–639.

Economic Commission for Europe (2006) High-level meeting on transport, environment and health - the PEP Steering Committee. (Fourth session, 10–11 April 2006, agenda item 4.A (c)) *Transport related health impacts and their costs and benefits with particular focus on children. Development of a 'toolbox' for action on transport, environment and health*. ECE/AC.21/2006/5 and EUR/06/THEPEPST/5.

Erath, AL, T Maheshwari, M Joos, J Kupferschmid and MAB Eggermond (2016) Visualising transport futures: the potential of integrating procedural 3D modelling and traffic micro simulation in virtual reality applications. *Arbeitsberichte Verkehrs – Und Raumplanung 1185*: 1–23.

Fosgerau, M (2016) The valuation of travel time variability. *International Transport Forum discussion paper 2016-04*.

Fosgerau, M, K Hjorth, S Vincent Lyk-Jensen and J Marott (2007) *The Danish value of time study: transfer pricing.* The Danish Transport Research Institute.

Galesic, M and M Bosnjak (2009) Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly 73*< no.2: 349–360.

Genter, JA, S Donovan, B Petrenas and H Badland (2008) Valuing the health benefits of active transport modes. *NZ Transport Agency research report 359*. 72pp.

Gravitas (2017) *Non-market transport impacts: choice modelling pilot questionnaire test (beta test)*. Gravitas, October 2017. p11.

Green, C and S Tunstall (1999) A psychological perspective. In *Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU, and developing countries.* I Bateman and KG Willis (Eds). Oxford, England: Oxford University Press.

Green, PE and V Srinivasan (1990) Conjoint analysis in marketing: new developments with implications for research and practice, *Journal of Marketing 54*, no.1: 3–19.

Groothuis, PA, JD Groothuis and JC Whitehead (2007) The willingness to pay to remove billboards and improve scenic amenities. *Journal of Environmental Management 85*, no.4: 1094–1100.

Groves, RM, FJ Fowler Jr, MP Couper, JM Lepkowski, E Singer and R Tourangeau (2004) *Survey methodology*. Hoboken, NJ: Wiley.

Guria, J (1990) *Estimation of social costs of traffic accidents and injuries*. Wellington: Ministry of Transport.

Guria, J, W Jones, MW Jones-Lee, M Keall, J Leung and G Loomes (2003) The New Zealand values of statistical life and of the prevention of injuries. Draft Report to New Zealand Land Transport Safety Authority.

Haab, TC, MG Interis, DR Petrolia and JC Whitehead (2013) From hopeless to curious? Thoughts on Hausman's 'dubious to hopeless' critique of contingent valuation. *Working papers 13–07*. Department of Economics, Appalachian State University.

Hartmann, A and FYY Ling (2016) Value creation of road infrastructure networks: a structural equation approach. *Journal of Traffic and Transportation Engineering 3*, no.1: 28–36.

Hatton MacDonald, D, JM Rose, H Lease and D Cox (2016) Recycled wastewater and product choice: Does it make a difference if and when you taste it? *Food Quality and Preference 48*: 283–292.

Hausman, J (2012) Contingent valuation: from dubious to hopeless*. The Journal of Economic Perspectives 26,* no.4: 43–56.

Hensher, DA (2001a) How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics 21*: 861–878.

Hensher, DA (2001b) Measurement of the valuation of travel time savings. *Journal of Transport Economics and Policy 35*, no.1: 71–98.

Hensher, DA (2004) Accounting for stated choice design dimensionality in willingness to pay for travel time savings, *Journal of Transport Economics and Policy 38*: 425–446.

Hensher, DA (2006) How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics 21*: 861–878.

Hensher, DA and WH Greene (2003) The mixed logit model: the state of practice. *Transportation 30*, no.2: 133–176.

Hensher, DA and C Sullivan (2003) Willingness to pay for road curviness and road type. *Transportation Research Part D: Transport and Environment 8*, no.2: 139–155.

Hensher, DA, JM Rose and WH Greene (2005*) Applied choice analysis: a primer*. New York, NY: Cambridge University Press.

Hensher, DA, JM Rose and WH Greene (2015) *Applied choice analysis*. 2nd ed. New York, NY: Cambridge University Press.

Hensher, DA, JM Rose, J de Dios Ortúzar and LI Rizzi (2009) Estimating the willingness-to pay and value of risk reduction for car occupants in the road environment. Institute of Transport and Logistics Studies *Working paper ITLS-WP-09-03*. University of Sydney.

The Hensher Group Pty Ltd (2007) *RTA NSW road safety project. Estimating the willingness-to-pay for road safety improvements*. Input to PwC Report for Roads and Traffic Authority of NSW.

Hess, S and JM Rose (2009) Some lessons in stated choice survey design. In *Vol 9 of European Transport Conference*.

Hess, S, M Bierlaire and JW Polak (2005) Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice 39*, no.2: 221–236.

HM Treasury (2003) *The green book: appraisal and evaluation in central government*. London: HM Treasury.

Holbrook, AL, MC Green and J Krosnick (2003) Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparison of respondent satisficing and social desirability response bias. *Public Opinion Quarterly 67*: 79–125.

House of Lords (2006) *Select committee on economic affairs fifth report.* Government policy in practice: valuation of safety.

Huber, F, A Hermann and A Gustafsson (2003) On the influence of the evaluation methods in conjoint design – some empirical results. Pp209–235 in *Conjoint measurement: methods and applications.* A Gustafsson, A Hermann and F Huber (Eds). New York, NY: Springer Verlag.

Hyman, W (2004) Planning and administration: maintenance guide for customer driven benchmarking of maintenance activities. *National Cooperative Highway Research Program (NCHRP) report 511*. Transportation Research Board.

Ian Wallis Associates Ltd (2014) Car passenger valuations of quantity and quality of time savings. *NZ Transport Agency research report 551*. 64pp.

Interagency Working Group on Social Cost of Carbon, United States Government (2015) *Technical support document: technical update of the social cost of carbon for regulatory impact analysis – under Executive Order 12866*. 2015 update.

Jäckle, A, C Roberts and P Lynn (2010) Assessing the effect of data collection mode on measurement. *International Statistical Review 78*, no.1: 3–20.

Johannesson, M and P-O Johansson (1997) Quality of life and the WTP for an increased life expectancy at an advanced age. *Journal of Public Economics 65*, no.2: 219–228.

Johnson, FR, E Lancsar, D Marshall, V Kilambi, A Mühlbacher, DA Regier and JF Bridges (2013) Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health 16*, no.1: 3–13.

Jones-Lee, M and M Spackman (2013) The development of road and rail transport safety valuation in the United Kingdom. *Research in Transportation Economics 43*: 23–40.

Kahlmeier, S, P Kelly, C Foster, T Götschi, N Cavill, H Dinsdale, J Woodcock, C Schweizer, H Rutter, C Lieb, P Oja and F Racioppi (2014) *Health economic assessment tools (HEAT) for walking and cycling. Methodology and user guide.* 2014 update. Economic Assessment of Transport Infrastructure and Policies. World Health Organisation.

Kamakura, W and GA Russell (1989) A probabilistic choice model for market segmentation and elasticity structuring. *Journal of Marketing Research* 26, no.4: 379–90.

Kerr, GN and BMH Sharp (2010) Choice experiment adaptive design benefits: a case study. *Australian Journal of Agricultural and Resource Economics 54*: 407–420.

Kim, S and TC Haab (2004) Revisiting bid design issues in contingent valuation. *Selected paper prepared for presentation at the American Agricultural Economics Association Annual Meeting*, Denver, Colorado, 1–4 August 2004.

Klojgaard, ME, M Bech and R Soggard (2012) Designing a stated choice experiment: the value of a qualitative process. *Journal of Choice Modelling 5*, no.2: 1–18.

Kreuter, F, S Presser and R Tourangeau (2008) Social desirability bias on CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly 72*: no.5: 847–865.

Krosnick, J (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology 5*: 213–236.

Kuschel, G, J Metcalfe, E Wilton, J Guria, S Hales, K Rolfe and A Woodward (2012) *Updated health and air pollution in New Zealand study. Volume 1: summary report.* Prepared for Health Research Council of New Zealand, Ministry of Transport, Ministry for the Environment and New Zealand Transport Agency

Land Transport NZ (2004) *A New Zealand guide to the treatment of crash locations. A companion document to the Austroads Guide to traffic engineering practice. Part 4. Treatment of crash locations*. Wellington: Land Transport NZ.

Lanscar, E and JJ Louviere (2008) Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics 26*: 661–667.

Lanscar, E and E Savage (2004) Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory, *Health Economics 13*, no.9: 901–907.

Lee, PSL (2012) *The application of choice modelling to determine the values of non-market goods and services: a national park case study.* Massey University: New Zealand.

Leung, J (2009) *Understanding transport costs and charges. Phase 2 – value of statistical life: a meta analysis. Is the current value of safety for New Zealand too low?* Wellington: Ministry of Transport.

Li, Z, DA Hensher and J Rose (2010) Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Institute of Transport and Logistics Studies working paper ITLS-WP-10-05*.

Lindhjem, H and S Navrud (2011) Using internet in stated preference surveys: a review of survey modes. *International Review of Environmental and Resource Economics 5*: 309–351.

Loomis, J (2011) What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys 25,* no.2: 363–370.

Louviere, JJ and HJP Timmermans (1990) Hierarchical information integration applied to residential choice behaviour. *Geographical Analysis 22*, no.2: 127–145.

Louviere, JJ, TN Flynn and AAJ Marley (2015) *Best-worst scaling: theory, methods and applications.* Cambridge, UK: Cambridge University Press.

Louviere, JJ, DA Hensher and JD Swait (2000) *Stated choice methods: analysis and application*. Cambridge, UK: Cambridge Press.

Lusk, JL and TC Schroeder (2004) Are choice experiments incentive compatible? A test with quality differential beef steaks. *American Journal of Agricultural Economics 5*: no.1: 467–482.

Mackie, P (2010) Cost-benefit analysis in transport – a UK perspective. In *OECD/International Transport Forum Joint Transport Research Centre Round Table*, Mexico. 21–22 October 2010.

Mahieu, P, P Riera and M Giergiczny (2012) The influence of cheap talk on willingness-to-pay ranges: some empirical evidence from a contingent valuation study. *Journal of Environmental Planning and Management 55*, no.6: 753–763.

Manfreda, KL, M Bosniak, J Berzelak, I Haas and V Vehovar (2008) Web surveys versus other survey modes: a meta-analysis comparing response rates. *International Journal of Market Research 50*, no.1: 79–104.

Mangham, LJ, K Hanson and B McPake (2009) How to do (or not to do)... Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning 24*, no.2: 151–158.

Mark, TL, and J Swait (2004) Using stated preference and revealed preference modelling to evaluate prescribing decisions. *Health Economics 13*: 563–573.

McFadden, D (1974) Conditional logit analysis of qualitative choice behaviour. Pp105–142 *Frontiers of econometrics*. P Zarembka (Ed). New York: Academic Press.

McMahon, K (1994) *Valuation of road accidents*. UK: Department for Transport.

McMahon, K and S Dahdah (2008) *The true cost of road crashes. Valuing life and the cost of a serious injury.* International Road Assessment Programme.

McNamara, JM, PC Trimmer and AI Houston (2014) Natural selection can favour 'irrational' behaviour. *Biology Letters 10*, no.1: 1–5.

Meade, R and L Cheung (2016) *GPS user preferences exploratory work*. Report for the New Zealand Ministry of Transport.

Metz, D (2008) The myth of travel time saving. *Transport Reviews 28*, no.3: 321–336.

Miller, TR (2000) Variations between countries in values of statistical life. *Journal of Transport Economics and Policy 34*, no.2: 169–188.

Miller, TR and J Guria (1991) *The value of statistical life in New Zealand: market research on road safety*. Wellington: Ministry of Transport.

Ministry of Transport (2010) *The social cost of road crashes and injuries: June 2010 update*.

Ministry of Transport (2016a) *Motor vehicle crashes in New Zealand 2015.* Section 1 Historical (Excel spreadsheet). Accessed 6 May 2018. www.transport.govt.nz/research/roadcrashstatistics/motorvehiclecrashesinnewzealand/motor-vehicle-crashes-in-new-zealand-2015/

Ministry of Transport (2016b) *Social cost of road crashes and injuries: 2015 update*. Accessed 6 May 2018. www.transport.govt.nz/resources/road-safety-resources/roadcrashstatistics/social-cost-of-road-crashes-and-injuries/

Mishan, EJ (1982) *Cost-benefit analysis: an informal introduction*. 3rd ed. London: George Allen & Unwin. 447pp.

Mitani, Y and NE Flores (2014) Hypothetical bias reconsidered: payment and provision uncertainties in a threshold provision mechanism. *Environmental Resource Economics 59*: 433–454.

Mitchell, RC and RT Carson (1989) *Using surveys to value public goods: the contingent valuation method*. Washington, DC: Resources for the Future.

Moller, SI (2012) *The economic value of environmental amenities and restoration for rural land in New Zealand*. Dunedin, New Zealand: Ecosystems Consultants Ltd.

Moore, MJ and WK Viscusi (1988) Doubling the estimated value of life: results using new occupational fatality data. *Journal of Policy Analysis and Management 7*, no.3: 476–490.

Moran, MJ and C Monje (2016) *Guidance on treatment of the economic value of a statistical life in US Department of Transportation analyses – 2016 adjustment*. Memorandum to Secretarial Officers and Modal Administrators. US Department of Transportation.

Morgan, MG and M Cropper (2007) SAB advisory on EPA’s issues in valuing mortality risk reduction. Letter to Stephen L Johnson, Administrator EPA. EPA-SAB-08-001.

Murphy, JJ, PG Allen, TH Stevens and D Weatherhead (2005) A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics 30*, no.3: 313–325.

National Research Council (2008) *Estimating mortality risk reduction and economic benefits from controlling ozone air pollution*. Committee on Estimating Mortality Risk Reduction Benefits from Decreasing Tropospheric Ozone Exposure. National Academies Press. Washington DC.

Naude, C, D Tsolakis, F Tan and T Makwasha (2015) Social costs of road crashes in Australia: the case for willingness-to-pay values for road safety. *Austroads Publication AP-R438-15.*

Nicolau, JL and FJ Más (2005) The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management 27,* no.5: 982–996.

Niroomand, N and GP Jenkins (2016) Estimating the value of life, injury, and travel time saved using a stated preference framework*. Accident Analysis & Prevention 91*: 216–225.

NZ Productivity Commission (2014) *Regulatory institutions and practices*. Wellington.

NZ Transport Agency (2015) *Statement of Intent 2015–19*. Wellington: NZ Transport Agency.

NZ Transport Agency (2016) *Economic evaluation manual*. Wellington: NZ Transport Agency.

OECD (2012) *Mortality risk valuation in environment, health and transport policies*. OECD Publishing.

O'Fallon, C and I Wallis (2012) A wider look at how travellers value the quality and quantity of travel time. *NZ Transport Agency research report 469*. 128pp.

Ohler T, A Li, JJ Louviere and J Swait (2000) Attribute range effects in binary response tasks, *Marketing Letters 11*: 249–260.

Olsson, C (2002) Motorists' evaluation of road maintenance management. Stockholm: Division of Systems Analysis and Economics Unit for Transport and Location Analysis. Royal Institute of Technology.

Oxera and Mott MacDonald (2003) *The scheduling of consumer activities: work trips*. Report to The Strategic Rail Authority.

Patterson, TL, WJ Frith and MW Small (2000) *Down with speed: a review of the literature, and the impact of speed on New Zealanders*. Wellington: Accident Compensation Corporation and Land Transport Safety Authority.

Pieters, R and L Warlop (1999) Visual attention during brand choice: the impact of time pressure and task motivation. *International Journal of Research in Marketing 16*: 1–16.

Prideaux, B (1999) The role of the transport system in destination development. *Tourism Management 21*, no.1: 53–63.

Quan, W, J Rose, A Collins and M Bliemer (2011) A comparison of algorithms for generating efficient choice. *Working paper ITLS-WP-11-19*. Sydney: University of Sydney. Institute of Transport and Logisitics Studies.

Raghothama, J and S Meijer (2013) A review of gaming simulation in transportation. Pp237–244 in *Frontiers in gaming simulation.* SA Meijer and R Smeds (Eds). Cham, Switzerland: Springer.

Ramjerdy, F, L Rand, IA Saetermo and K Saelensminde (1997) The Norwegian value of time study, part I and part II. *TØI report 379/1997*. Oslo.

Rao, VR (2014) Theory and design of conjoint studies (ratings based methods). Chapter 2, pp37–78 in *Applied conjoint analysis*. VR Rao. Berlin Heidelberg: Springer.

Revelt, D and K Train (1998) Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics 80*, no.4: 647–657.

Richardson, A (2003) Some evidence of travelers with zero value-of time. *Transportation Research Record 1854*: 107–113.

Risbey T, H de Silva and A Tong (2007) *Road crash cost estimation: a proposal incorporating a decade of conceptual and empirical developments*. Canberra, Australia: Bureau of Transport and Regional Economics.

Rizzi, LI and J de Dios Ortúzar (2003) Stated preference in the valuation of interurban road safety. *Accident Analysis & Prevention 35*, no.1: 9–22.

Rizzi, LI and J de Dios Ortuzar (2006) Road safety valuation under a stated choice framework. *Journal of Transport Economics and Policy 40*, no.1: 69–94.

Robinson, LA (2007) Policy monitor: how US government agencies value mortality risk reductions. *Review of Environmental Economics and Policy 1*, no.2: 283–299.

Rolfe, J and J Bennett (2009). The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics 68*, no.4: 1140–1148.

Rose, JM and MC Bliemer (2009) Constructing efficient stated choice experimental designs. *Transport Reviews*, 29(5):587-617.

Rose, JM and MC Bliemer (2012) Sample optimality in the design of stated choice experiments. Travel behaviour research in the evolving world. Pp119–45 in *Selected papers from the 12th International Conference on Travel Behaviour Research*. R Pendyala and C Bhat (Eds), India.

Rose, JM and DA Hensher (2006) Accounting for individual specific non-availability of alternatives in respondent's choice sets in the construction of stated choice experiments. In *Survey methods*. PR Stopher and C Stecher (Eds). Oxford: Elsevier Science.

Rose, JM, DA Hensher, S Caussade, J de Dios Ortuzar and J Rong-Chang (2009) Identifying differences in preferences due to dimensionality in stated choice experiments: a cross cultural analysis. *Journal of Transport Geography 17*, no.1: 21–29.

Rouwendal, J, A de Blaeij, P Rietveld and E Verhoef (2010) The information content of a stated choice experiment: A new method and its application to the value of a statistical life. *Transportation Research Part B 44*: 136–151.

Sælensminde, K (2004) Cost-benefit analyses of walking and cycling track networks taking into account insecurity, health effects and external costs of motorized traffic. *Transportation Research Part A: Policy and Practice 38*, no.8: 593–606.

Sagoff, M (1988) *The economy of the earth*. Cambridge: Cambridge University Press.

Salomon, I and P Mokhtarian (1998) What happens when mobility-inclined market segments face accessibility-enhancing policies? *Transportation Research D 3*, no.3: 129–140.

Sanko, N (2001) *Guidelines for stated preference experiment design*. Master of Business Administration, Ecole Nationale des Ponts et Chaussées.

Sawtooth Software (nd) *Lighthouse Studio: ACA*. Accessed 6 May 2018. www.sawtoothsoftware.com/products/conjoint-choice-analysis/aca

Scarpa, R and JM Rose (2008) Designs efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics 52*, no.3: 253–282.

Schroeder, HW and J Louviere (1999) Stated choice models for predicting the impact of user fees at public recreation sites. *Journal of Leisure Research 31*, no.3: 300–324.

Significance, VU University and John Bates Services (2012) *Values of time and reliability in passenger and freight transport in The Netherlands*. Report for the Ministry of Infrastructure and the Environment.

Simon, HA (1955) A behavioural model of rational choice. *The Quarterly Journal of Economics 69*, no.1: 99–118.

Small, KA (1982) The scheduling of consumer activities: work trips. *The American Economic Review 72*, no.3: 467–479.

Small, KA, RB Noland, X Chu and D Lewis (1999) Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation. *NCHRP report 431*, Transportation Research Board, National Research Council.

Sommer, H, R Seethaler, O Chanel, M Herry, S Masson and J-C Vergnaud (1999) *Health costs due to road traffic-related air pollution. An impact assessment project of Austria, France and Switzerland*. Prepared for the WHO Ministerial Conference on Environment and Health London June 1999.

Stüttgen, P, P Boatwright and RT Monroe (2011) A satisficing choice model. *Marketing Science 31*, no.6: 878–899.

Swait, J and W Adamowicz (2001) The influence of task complexity on consumer choice: a latent class model of decision strategy switching. *Journal of Consumer Research 28*, no.1: 135–148.

Szolnoki, G and D Hoffmann (2012) Online, face-to-face and telephone surveys – comparing different sampling methods in wine consumer research. *Wine Economics and Policy 2*, no.2: 57–66.

Telser, H and P Zweifel (2002) Measuring willingness-to-pay for risk reduction: an application of conjoint analysis. *Health Economics 11*: 129–139.

Telser, H and P Zweifel (2007) Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics 39*, no.1: 69–78.

Thomas, PJ and GJ Vaughan (2015) Testing the validity of the 'value of a prevented fatality' (VPF) used to assess UK safety measures. *Process Safety and Environmental Protection 94*: 239–261.

Thurstone, L (1931) The indifference function. *Journal of Social Psychology 2*, no.2: 139–167.

Tilahun, NY and DM Levinson (2010) A moment of time: reliability in route choice using stated preference. *Journal of Intelligent Transportation Systems 14*, no.3: 179–187.

Tourangeau, R and T Yan (2007) Sensitive questions in surveys. *Psychological Bulletin 133*: 859–883.

Train, KE (2002) *Discrete choice methods with simulation* (vol. 8). Cambridge: Cambridge University Press.

Train, KE, DL McFadden and AA Goett (1987) Consumer attitudes and voluntary rate schedules for public utilities. *The Review of Economics and Statistics 6*9, no.3: 383–391.

Trottenberg, P and RS Rivkin (2013) *Guidance on treatment of the economic value of a statistical life in US Department of Transportation analyses: revised departmental guidance 2013: treatment of the value of preventing fatalities and injuries in preparing economic analyses*. US Department of Transportation.

Tseng, YY and ET Verhoef (2008) Value of time by time of day: a stated-preference study. *Transportation Research Part B: Methodological 42*, no.7: 607–618.

Tsolakis, D, C Naude, F Tan, C Evans, T Makwasha and J Shackleton (2011) Updating Austroads RUE unit values and related methodologies. *Austroads research report AP-R373/11*.

Tsuge, T, A Kishimoto and K Takeuchi (2005) A choice experiment approach to the valuation of mortality. *Journal of Risk and Uncertainty 31*, no.1: 73–95.

Turner, RK (1999) The place of economic values in environmental valuation. Pp17–41 in *Valuing environmental preferences.* IJ Bateman and KG Willis (Eds). New York: NY: Oxford.

Tversky, A (1972) Elimination by aspects: a theory of choice. *Psychological Review 79*, no.4: 281–299.

US Environmental Protection Agency (2010) *Valuing mortality risk reductions for environmental policy: a white paper* (SAB review draft). Prepared by the National Center for Environmental Economics for consultation with the Science Advisory Board – Environmental Economics Advisory Committee.

Van der Pol, M, G Currie, S Kromm and M Ryan (2014) Specification of the utility function in discrete choice experiments. *Value in Health 17*, no.2: 297–301.

van der Waerden, P, A Borgers and H Timmermans (2004) The effects of attribute level definition on stated choice behaviour. In *Proceedings of the 7th International Conference on Travel Survey Methods*.

Vanniyasingam, T, C Cunningham, G Foster and L Thabane (2016) Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. *BMJ Open 6*, no.7: 1–10.

Veisten, K, S Flügel, LI Rizzi, J de Dios Ortúzar and R Elvik (2013) Valuing casualty risk reductions from estimated baseline risk. *Research in Transportation Economics 43*, no.1: 50–61.

Viscusi, WK (2005) The value of life. *John M Olin Center for Law, Economics and Business discussion paper no.517.*

Viscusi, WK and JE Aldy (2003) The value of a statistical life: a critical review of market estimates throughout the world. *National Bureau of Economic Research working paper 9487.*

Von Neumann, J and O Morgenstern (1947) *Theory of games and economic behaviour.* Princeton University Press, 2nd ed.

Wardman M, PJ Mackie and A Gillies-Smith (2014) Valuing systemic transport resilience: methods and evidence. Pp18–41 in *Economic evaluation of systems of infrastructure provision: concepts, approaches, methods.* A Brown and M Robertson (Eds).

Wallis, I, K Rupp and R Alban (2015) Travel time saving assessment. *NZ Transport Agency research report 570*. 157pp.

Wang, D, L Jiuqun and HJP Timmermans (2001) Reducing respondent burden, information processing and incomprehensibility in stated preference surveys: principles and properties of paired conjoint analysis. *Transportation Research Record 1768*: 71–78.

Windle, J and J Rolfe (2011) Comparing response from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis and Policy 41*, no.1: 83–97.

Wren, J and K Barrell (2010) T*he costs of injury in New Zealand and methods for prioritising resource allocation: New Zealand injury prevention strategy*. Wellington: New Zealand Injury Secretariat, ACC.

Wu, Y and G Fleming (2012) *Work trip route choice survey*. Poster presentation at Southeast Florida FSUTMS Users Group Meeting. Ft. Lauderdale, Florida.

Yáñez, MF and J Ortúzar (2010) Modelling choice in a changing environment: assessing the shock effects of a new transport system. Pp445–460 in *Choice modelling: the state-of-the-art and the state-of-practice.* S Hess and A Daly (Eds). Bingly, United Kingdom: Emerald Group Publishing.

Yusoff, MFM, NA Mohamad, NZ Abidin, NGM Nor and H Salleh (2013) The value of statistical life in fatal injury among drivers and riders in Malaysia: conjoint analysis method. *MIROS internal report no.130*.

Appendix A: Glossary

AAAM	Association for the Advancement of Automotive Medicine (US)
ACA	adaptive conjoint analysis
ACC	Accident Compensation Corporation
aBIC	adjusted BIC
AIC	Akaike information criterion
AIC3	Bozdogan AIC
AIS	abbreviated injury scale
ASC	alternative specific constant
BIC	Bayesian information criterion
BITRE	Bureau of Infrastructure, Transport and Regional Economics (Australia)
BTE	Bureau of Transport Economics (Australia)
BTRE	Bureau of Transport and Regional Economics (Australia)
CAIC	consistent AIC
CATI	computer-assisted telephone interviewing
CBA	cost-benefit analysis
CM	choice model/modelling
COMEAP	Committee on the Medical Effects of Air Pollution (UK)
CPI	consumer price index
CV	contingent valuation
DOT	Department of Transportation (US)
EEM	*Economic evaluation manual* (NZ Transport Agency 2016)
GDP	gross domestic product
GPS	global positioning system
HCA	human capital approach
HM	Her Majesty's
HTS	Household Travel Survey
HW	hedonic wage
MLE	maximum likelihood estimation
M	million
MNL	multinomial logit
MoT	Ministry of Transport (New Zealand)

NZIER	New Zealand Institute of Economic Research
OECD	Organisation for Economic Co-operation and Development
PAT	preferred arrival time
PM&C	Department of Prime Minister & Cabinet (Australia)
RP	revealed preference
RPL	random parameters logit (model)
RTA	Roads and Traffic Authority (Australia)
SEM	standard error of the mean
SP	stated preference
VKT	vehicle kilometres travelled
VoLY	value of a life year
VOR	value of reliability
VoSL	value of a statistical life
VOT	value of time
VSL	value of a statistical life
VTTS	value of travel time savings
WHO	World Health Organisation
WTA	willingness to accept
WTP	willingness to pay

Appendix B: Efficient experimental design

Below we set out the experimental design. This design incorporates two sample splits to allow identification of VSL and the value of injuries by utilisation of two different rates of injuries relative to fatalities. There are three blocks in each scenario, yielding six different choice sets (labelled Person 1 through Person 6). The blocking was identified as part of the experimental design. Each individual participating in the study will face 10 choice scenarios, up from eight in the beta-test. This design maintains attribute balance. Participants will be allocated a questionnaire version ('Person I') on rotation to ensure equal numbers of each version in the final sample.

B1 Version 1

Person 1		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	240	240	210	180	240	180	210	240	180	210	20	40	30	40	20	40	40	20	30	20
Lateness	Minutes	30	30	15	40	40	30	30	30	40	15	5	10	10	10	20	5	10	20	10	10
Heavy traffic	Minutes	60	0	30	0	0	60	30	30	30	30	0	20	10	10	20	0	10	10	10	10
Risk of Death	per bVKT	7.5	2.5	2.5	2.5	5	5	2.5	7.5	7.5	7.5	6	6	10	2	6	6	2	10	2	10
Percent	relative to NZ mean	139%	46%	46%	46%	93%	93%	46%	139%	139%	139%	111%	111%	185%	37%	111%	111%	37%	185%	37%	185%
Serious Injury	per bVKT	38	13	13	13	25	25	13	38	38	38	30	30	50	10	30	30	10	50	10	50
Minor injury	per bVKT	263	88	88	88	175	175	88	263	263	263	210	210	350	70	210	210	70	350	70	350
Cost	$	70	80	60	60	60	70	80	70	70	70	9	3	3	6	6	9	3	3	6	3

Person 2		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	180	210	210	180	240	210	180	240	180	210	30	30	40	20	20	30	40	20	40	20
Lateness	Minutes	15	40	40	15	15	15	40	30	15	40	20	10	5	20	20	5	5	20	10	5
Heavy traffic	Minutes	60	30	30	30	60	0	0	60	0	60	20	10	0	20	10	0	0	20	20	0
Risk of Death	per bVKT	2.5	7.5	7.5	7.5	2.5	2.5	5	5	7.5	2.5	10	2	10	10	2	2	10	2	6	6
Percent	relative to NZ mean	46%	139%	139%	139%	46%	46%	93%	93%	139%	46%	185%	37%	185%	185%	37%	37%	185%	37%	111%	111%
Serious Injury	per bVKT	13	38	38	38	13	13	25	25	38	13	50	10	50	50	10	10	50	10	30	30
Minor injury	per bVKT	88	263	263	263	88	88	175	175	263	88	350	70	350	350	70	70	350	70	210	210
Cost	$	80	60	60	70	70	80	80	70	80	80	6	6	9	9	9	9	9	9	3	9

Person 3		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	210	180	180	210	240	180	210	240	240	240	40	30	20	40	20	40	30	30	30	30
Lateness	Minutes	30	15	30	30	15	40	30	15	40	40	5	20	5	20	20	5	10	10	20	5
Heavy traffic	Minutes	60	0	0	60	0	60	30	30	60	0	20	0	20	0	0	20	10	10	0	20
Risk of Death	per bVKT	2.5	7.5	5	5	5	5	5	2.5	7.5	5	2	2	6	6	6	6	2	10	10	10
Percent	relative to NZ mean	46%	139%	93%	93%	93%	93%	93%	46%	139%	93%	37%	37%	111%	111%	111%	111%	37%	185%	185%	185%
Serious Injury	per bVKT	13	38	25	25	25	25	25	13	38	25	10	10	30	30	30	30	10	50	50	50
Minor injury	per bVKT	88	263	175	175	175	175	175	88	263	175	70	70	210	210	210	210	70	350	350	350
Cost	$	80	80	70	60	60	60	60	60	70	80	6	6	9	3	3	3	3	6	6	6

B2 Version 2

Person 4		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	240	240	210	180	240	180	210	240	180	210	20	40	30	40	20	40	40	20	30	20
Lateness	Minutes	30	30	15	40	40	30	30	30	40	15	5	10	10	10	20	5	10	20	10	10
Heavy traffic	Minutes	60	0	30	0	0	60	30	30	30	30	0	20	10	10	20	0	10	10	10	10
Risk of Death	per bVKT	7.5	2.5	2.5	2.5	5	5	2.5	7.5	7.5	7.5	6	6	10	2	6	6	2	10	2	10
Percent	relative to NZ mean	139%	46%	46%	46%	93%	93%	46%	139%	139%	139%	111%	111%	185%	37%	111%	111%	37%	185%	37%	185%
Serious Injury	per bVKT	68	23	23	23	45	45	23	68	68	68	54	54	90	18	54	54	18	90	18	90
Minor injury	per bVKT	473	158	158	158	315	315	158	473	473	473	378	378	630	126	378	378	126	630	126	630
Cost	$	70	80	60	60	60	70	80	70	70	70	9	3	3	6	6	9	3	3	6	3

Person 5		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	180	210	210	180	240	210	180	240	180	210	30	30	40	20	20	30	40	20	40	20
Lateness	Minutes	15	40	40	15	15	15	40	30	15	40	20	10	5	20	20	5	5	20	10	5
Heavy traffic	Minutes	60	30	30	30	60	0	0	60	0	60	20	10	0	20	10	0	0	20	20	0
Risk of Death	per bVKT	2.5	7.5	7.5	7.5	2.5	2.5	5	5	7.5	2.5	10	2	10	10	2	2	10	2	6	6
Percent	relative to NZ mean	46%	139%	139%	139%	46%	46%	93%	93%	139%	46%	185%	37%	185%	185%	37%	37%	185%	37%	111%	111%
Serious Injury	per bVKT	23	68	68	68	23	23	45	45	68	23	90	18	90	90	18	18	90	18	54	54
Minor injury	per bVKT	158	473	473	473	158	158	315	315	473	158	630	126	630	630	126	126	630	126	378	378
Cost	$	80	60	60	70	70	80	80	70	80	80	6	6	9	9	9	9	9	9	3	9

Person 6		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10	
		Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2	Route 1	Route 2
Travel time	Minutes	210	180	180	210	240	180	210	240	240	240	40	30	20	40	20	40	30	30	30	30
Lateness	Minutes	30	15	30	30	15	40	30	15	40	40	5	20	5	20	20	5	10	10	20	5
Heavy traffic	Minutes	60	0	0	60	0	60	30	30	60	0	20	0	20	0	0	20	10	10	0	20
Risk of Death	per bVKT	2.5	7.5	5	5	5	5	5	2.5	7.5	5	2	2	6	6	6	6	2	10	10	10
Percent	relative to NZ mean	46%	139%	93%	93%	93%	93%	93%	46%	139%	93%	37%	37%	111%	111%	111%	111%	37%	185%	185%	185%
Serious Injury	per bVKT	23	68	45	45	45	45	45	23	68	45	18	18	54	54	54	54	18	90	90	90
Minor injury	per bVKT	158	473	315	315	315	315	315	158	473	315	126	126	378	378	378	378	126	630	630	630
Cost	$	80	80	70	60	60	60	60	60	70	80	6	6	9	3	3	3	3	6	6	6

Appendix C: Survey

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Thank you for agreeing to take part in this survey.

The research is for the New Zealand Transport Agency, the government agency that aims to achieve an efficient, effective and safe land transport system.

The Transport Agency wants to know how to identify the kind of transport system users want, and this will help them to make decisions which balance safety, travel time and other outcomes.

The survey should take about 15 minutes. Your answers are confidential and anonymous and will only be used or reported in combination with other survey participants.

For more information about the survey, click **Here**

>>

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

(If clicked the button for more information in the introduction)

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an efficient, effective and safe land transport system in the public interest. Each year, the Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in the survey are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the Transport Agency. The material contained in the survey should not be construed in any way as policy adopted by the Transport Agency or indeed any agency of the NZ Government. The survey may, however, be used by NZ Government agencies as a reference in the development of policy.

If you have any further questions about the research, please contact Michael at Gravitas Research on 09 917 0961 or email michael@gravitas.co.nz

Click the forward arrow below to continue with the survey

<< >>

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

The following questions ask about the types of trips that you make in a car (or van, or ute etc) as the **driver**.

Q1 In the last **12 months**, how often have you made the following types of trips, as the driver?

Please select one per row

	6-7 days a week	5 days a week	3-4 days a week	1-2 days a week	About once every two weeks	About once a month	Less often than once a month	Never/ Not in the last 12 months
A local trip in your town, city or area for shopping, personal business (eg. banking, the gym) or social visit, which is not so urgent	○	○	○	○	○	○	○	○
A longer distance trip between city/town centres for recreation, or to visit family/friends, or for work purposes	○	○	○	○	○	○	○	○

<< >>

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q2 For the following trip[If 2x selected at Q1, "s"], please tell us how long you spent in the car/vehicle (excluding any stops you may have made)

We are only interested in one leg of your journey. If you made a return trip, only provide a time for one direction.

Please specify [If 2x selected at Q1, "for each."]

(If Q1 local trip is not = "never", ask):
For your most recent local trip in your town, city or area for shopping, personal business (eg banking, the gym) or social visit, which is not so urgent?
__________hours __________minutes

(If Q1 long distance trip is not = "never", ask):
For your most recent longer distance trip between city/town centres for recreation, or to visit family/friends, or for work purposes?

__________hours __________minutes

Note: a response is required for either the hours or minutes (cannot leave both blank)

<< >>

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q3 Which of the following best describes who pays for your fuel (petrol/diesel) and running costs?
Please select one only

- I pay for all of my own fuel
- Work pays for all of my fuel
- Someone else pays for all of my fuel
- A combination of the above (e.g. work pays for some trips and I pay for others)

<< >>

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

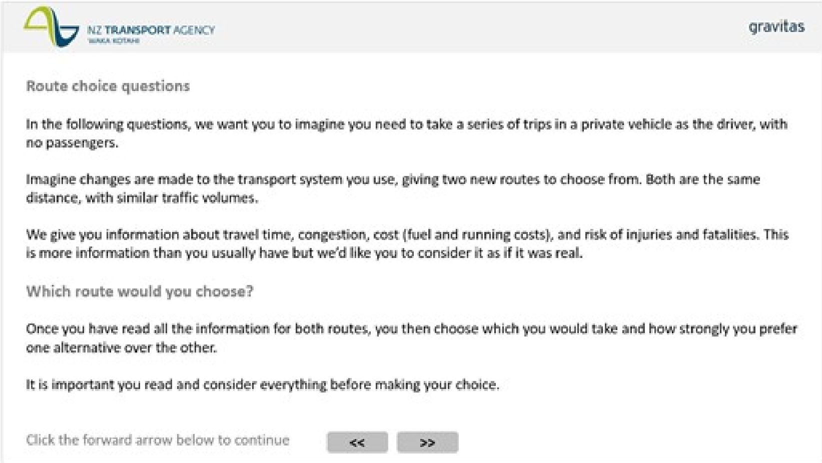
Q4 Which of the following best describes the area you live in?

If you move around or live in multiple areas, please select where you spend most of your time

Please select one only

- Rural area
- Town (e.g. Dargaville or Kaikoura)
- Smaller city (e.g. Gisborne, Invercargill, Nelson, New Plymouth, Palmerston N, Rotorua, Whangarei, Whanganui)
- Medium-sized city (e.g. Dunedin, Napier-Hastings, Tauranga)
- Larger city (e.g. Auckland, Christchurch, Hamilton, Wellington)

<< >>



NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Route choice questions

In the following questions, we want you to imagine you need to take a series of trips in a private vehicle as the driver, with no passengers.

Imagine changes are made to the transport system you use, giving two new routes to choose from. Both are the same distance, with similar traffic volumes.

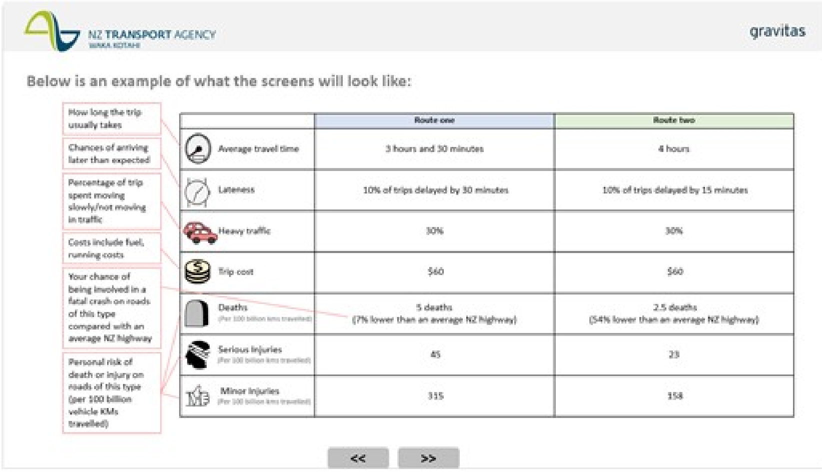
We give you information about travel time, congestion, cost (fuel and running costs), and risk of injuries and fatalities. This is more information than you usually have but we'd like you to consider it as if it was real.

Which route would you choose?

Once you have read all the information for both routes, you then choose which you would take and how strongly you prefer one alternative over the other.

It is important you read and consider everything before making your choice.

Click the forward arrow below to continue << >>

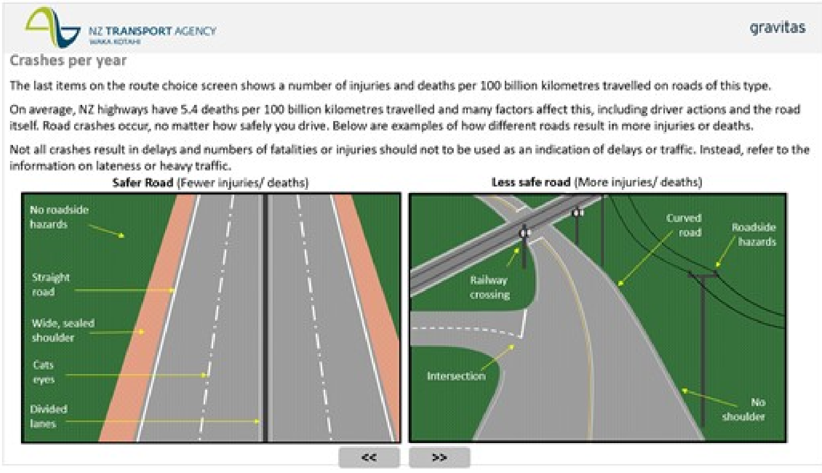


NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Below is an example of what the screens will look like:

	Route one	Route two
Average travel time	3 hours and 30 minutes	4 hours
Lateness	10% of trips delayed by 30 minutes	10% of trips delayed by 15 minutes
Heavy traffic	30%	30%
Trip cost	$60	$60
Deaths (Per 100 billion kms travelled)	5 deaths (7% lower than an average NZ highway)	2.5 deaths (54% lower than an average NZ highway)
Serious Injuries (Per 100 billion kms travelled)	45	23
Minor Injuries (Per 100 billion kms travelled)	315	158



NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Crashes per year

The last items on the route choice screen shows a number of injuries and deaths per 100 billion kilometres travelled on roads of this type.

On average, NZ highways have 5.4 deaths per 100 billion kilometres travelled and many factors affect this, including driver actions and the road itself. Road crashes occur, no matter how safely you drive. Below are examples of how different roads result in more injuries or deaths.

Not all crashes result in delays and numbers of fatalities or injuries should not to be used as an indication of delays or traffic. Instead, refer to the information on lateness or heavy traffic.

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

(Show this screen if Q3 = codes 2-4)

We understand that [If Q3=2, "your work" / if Q3=3, "someone else" / if Q3=4, "your work or someone else"] pays for [if Q3=2 or 3, "all of" / if Q3=4, "some of"] your fuel (petrol/diesel).

For the following, imagine you are **paying the trip cost yourself** and make your decision based on what you would do if you were paying.

<< >>

Ten games will be given, using a range of attribute levels.

Refer to the Attribute Levels excel document for more information on the levels used in each game.

The next screen is an example of what the games should look like.

The order of the 10 games will be randomised.

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Below is an example of how screens should look. Refer to the excel document for values displayed for each game

Out of the two alternatives shown, which one would you prefer to take?

Please select one only

	Route one	Route two
Average travel time	3 hours 15 minutes	3 hours 0 minutes
Lateness	10% of trips delayed by 40 minutes	10% of trips delayed by 45 minutes
Heavy traffic	20%	10%
Trip cost	$15	$20
Fatalities (Per 100 billion kms travelled)	2 deaths (50% higher than an average NZ highway)	1 death (10% lower than an average NZ highway)
Serious Injuries (Per 100 billion kms travelled)	10	2
Minor Injuries (Per 100 billion kms travelled)	20	5

○ **Strongly** prefer route one ○ **Moderately** prefer route one ○ **Slightly** prefer route one ○ **Slightly** prefer route two ○ **Moderately** prefer route two ○ **Strongly** prefer route two

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q6 Please select any items below that you did **not** consider when making your route choices

Please select all that apply

- Average travel time
- Lateness
- Heavy traffic
- Trip cost
- Deaths
- Serious injuries
- Minor injuries
- None of the above (exclusive)

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Thanks, that's all the route choice questions completed.

We just need a few more details about you to help us analyse the results, then we are done.

Before you move on to the next section, do you have any comments you would like to make about the choice questions or any of the decisions you have made?

Please specify **(Not required)**

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q7 How many vehicles of the types below are usually parked at your dwelling, or available to your household (but parked elsewhere)?

Please specify

Cars/Vans/Utes____________________
Motorcycles____________________
Motor scooters____________________
Trucks____________________

Note: a response is required for at least one vehicle type (cannot leave all blank)

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q8 Which driving licence type(s) do you currently hold?

Please select all that apply

- Car (full)
- Car (restricted)
- Car (learners)

- Motorcycle (full)
- Motorcycle (restricted)
- Motorcycle (learners)

- Heavy vehicle (full)
- Heavy vehicle (learners)

Only allow 1x car, 1x motorbike and 1x heavy vehicle licence type to be selected

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q9 For each injury type below, please indicate if you or someone you know have experienced a crash in a vehicle resulting in any of the following (irrespective of who was at fault)?

Please select all that apply.

	Personally suffered select all that apply	**In the car with someone else who suffered** select all that apply	**Someone you know but were not in the car with** select all that apply
No injury Crash resulting in no physical injuries	○	○	○
Minor injury Requiring treatment but not hospitalisation or follow up care e.g. cuts, bumps, whiplash	○	○	○
Major Injury Requiring hospitalisation e.g. broken bones, psychological distress, major bleeding or paralysis	○	○	○
Fatal injury Death within 30 days of the crash		○	○

- **None of the above** (I haven't been involved in any crash and don't know anyone else who has)

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q10 What is your year of birth?

Please select your year of birth from the drop-down list below

______________________ (Show drop down list of dates)

Q11 What is your gender?

Please select one only

- Male
- Female
- Other

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI
gravitas

Q12 Including yourself, how many adults (18 years or over) and children (younger than 18 years) live in your household?

Please specify for both below

Adults ______________
Children ______________

Must = at least 1 for either children or adults. Cannot leave both blank or enter 0 for both

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI
gravitas

Q13 What is the highest level of education you have achieved?
Please select one only

- High school (level 1-4 certificate)
- Diploma (level 5 and 6)
- Bachelor's Degree (level 7)
- Postgraduate Degree (Honours, Masters or Doctorate)
- Other (specify)______________

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI
gravitas

Q14 What best describes your current occupation?
If you have more than one occupation, which do you spend the most time doing?
Please select one only

- Student (full time)
- Student (part time)
- Employed (full time)
- Employed (part time)
- Employed (casually)
- Not working for pay
- Homemaker (full time)
- Volunteer worker (regular)
- Retired
- Unemployed (and seeking work)
- Other (specify)______________

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q15 What is your personal income (before tax)?
Please select one only

- No income
- $20,000 or less
- $20,001 to $30,000
- $30,001 to $50,000
- $50,001 to $70,000
- $70,001 to $100,000
- $100,001 or more

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Q16 Which type(s) of insurance do you have?
Please select all that apply

- Vehicle – Third party, fire and theft
- Vehicle – Comprehensive insurance
- Vehicle – Extended warranty
- Property – Home insurance
- Property – Contents insurance
- Income or mortgage protection
- Life insurance
- Pet insurance
- Private health insurance
- Other (specify)_______________
- No insurance (exclusive)

Next

NZ TRANSPORT AGENCY
WAKA KOTAHI

gravitas

Thank you for completing the survey.

If you have any questions about the research, please contact Michael from Gravitas Research on 09 9170961 or Michael@gravitas.co.nz

Please click on the 'Submit' button below to register your answers and close the survey.